DAILEY, GABRIELLE, P., Ph.D., RNA Viruses: The Strategies They Employ to Interfere with Host Cell Protein Expression and to Synthesize Viral Proteins, and a New Way to Identify Their RNA in Biological Samples. (2021)
Directed by Dr. Ethan Will Taylor. pp. 133

This dissertation is composed of the projects I worked on that were able to be made into scientific papers. My projects centered primarily around viral RNAs and how they interact with host cell mRNAs coding for selenoprotein, and if these effects were negated or enhanced by the addition of selenium to the environment. The first chapter of this dissertation is an introduction into selenoprotein synthesis and importance, viral implications in history, and viral survival strategies. My first author publication is the second chapter of this dissertation, and the one I spent the majority of my time on. It concerns predicted antisense interactions between conserved regions in the 3' untranslated region of Zika viral RNA and selenoprotein mRNA's in the host cell. Aside from the initial computational data for the antisense interactions between Zika (ZIKV) RNA and SePP1 and TXNRD1 host cell RNA species, along with the gel shift assay using oligonucleotides that matched the predicted sequences, all work in the first chapter of this dissertation was performed. This manuscript has been submitted to BBA Advances. The third chapter is the continuation of Dr. Lakmini Premadasa's work supporting the hypothesis that an antisense tethering region between TXNRD1 mRNA in the host cell and the 3' untranslated region downstream of the *nef* gene in HIV-1 could lead to a UGA stop codon readthrough at the 3' end of the *nef* gene resulting in a longer nef protein isoform. There were some issues reviewers posed before it could be published, and Dr. Premadasa was no longer around to be able to address the concerns, so my contributions to this manuscript involved a Western blot probe of GFP expression as a result of the proposed UGA stop codon readthrough event and if selenium had any effect on GFP expression. This manuscript preprint is published and the manuscript itself has been resubmitted to the American Journal of Pharmacotherapy and Pharmaceutical

Sciences. For the fourth chapter of this dissertation, I was lucky to get to collaborate with Dr. Adam Hall and Dr. Komal Sethi at the Wake Forest School of Medicine with a project regarding HIV-1 identification via Solid-State nanopore technology. I was responsible for synthesizing all RNA used for the experiments and was part of the HIV-1 and probe designs used in the experiments. I provided a write up of my RNA synthesis, purification, and quantification procedure for the paper. This manuscript has officially been approved for publication in ACS Nano.

RNA VIRUSES: THE STRATEGIES THEY EMPLOY TO INTERFERE WITH HOST CELL PROTEIN EXPRESSION AND TO SYNTHESIZE VIRAL PROTEINS, AND A NEW WAY TO IDENTIFY THEIR RNA IN BIOLOGICAL SAMPLES

by

Gabrielle Patricia Dailey

A DISSERTATION

Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro

2021

Approved by

Dr. Ethan Will Taylor
Committee Chair

DEDICATION

Dedicated to my husband for his relentless and sometimes questionable commitment to me.

Dedicated to my mother for a lifetime of love and support. Dedicated to me for not giving up.

Dedicated to my grandfather, who I wish was here.

APPROVAL PAGE

This dissertation written by Gabrielle Patricia Dailey has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair Dr. Ethan Will Taylor

Committee Members Dr. Nadja Cech

Dr. Norman Chiu

Dr. Jan Ruzicka

4/29/2021

Date of Acceptance by Committee

4/29/2021

Date of Final Oral Examination

AKNOWLEDGEMENTS

It was a privilege to be able to conduct my research at the University of North Carolina at Greensboro. It fills me with no end of pride that I got my advanced education at a university that has been a hallmark of equal education for so many women before me. I know where I am today is surrounded by giant shards of glass left on the ground from other women in STEM who managed to breakthrough and create a space for myself and other young woman scientists entering the field. I was lucky enough to have a very emotionally supportive mentor, Dr. Ethan Will Taylor, who embraced and supported me as an individual throughout this journey. I was given ample room to grow and manifest myself into an independent, confident, and persistent scientist under his tutelage.

I am grateful to all my committee members, Dr. Ethan Will Taylor, Dr. Nadja Cech, Dr. Norman Chiu, and Dr. Jan Ruzicka for the constant support they offer me; and for them being willing to sacrifice their valuable time to reading and evaluating this monster of a document. I'm grateful for every time I had a last-minute paperwork emergency and I was met with no judgement when I spam emailed them for help that I could have asked for weeks in advance. I am especially grateful to Dr. Nadja Cech, who has always expressed belief in me and advocated for me. She was part of the committee that granted me T32 funding, which freed up an enormous amount of my time and energy. I am grateful to the T32 community at UNCG for providing me with a support group that I didn't know I needed; a group which helped mold me into the scientist and communicator I get to be present day. She is someone I very much look up to.

I would like to specifically extend my gratitude to Dr. Jan Ruzicka, who opened up his lab space to me at the Fred Wilson School of Pharmacy at High Point University. He placed an inordinate

amount of trust on me and my abilities and allowed me to step into a leadership role for his lab. This research would not exist without his support, trust, and generosity. We have worked closely together for 4 years now and he has been my most consistent point of human contact in research.

I would very much like to thank my husband, family, and friends; all who despite evidence to the contrary have never once doubted me or my abilities. My mother is a constant reassurance and she never gets frustrated that I do need that reassurance constantly. She has been my supplier of unconditional love since before I can remember and red wine since I started graduate school. I am sincerely grateful to my husband, who has been living the graduate student stipend lifestyle with me and has supported me through it the entire way. Every time I thought I was putting our lives on hold by obtaining this degree, he was quick to remind me of what we were building together with this education. I always wanted a higher degree and would have sought one out regardless of me meeting my husband, but I doubt I would have retained as much sanity as I have without him by my side.

Finally, I'd like to extend my love and gratitude to my grandfather, who unfortunately passed from this world my second year at UNCG. He was the absolute loudest supporter of me in any room, and now every room feels a little more silent in his absence.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

1.1 Selenoproteins and their interest concerning viral/host interactions

Selenoproteins are a class of proteins important for various functions in the body. They are essential components in a variety of redox reactions in the cell and have cell maintenance/homeostasis functions, as well as play a key role in immune response [6, 7]. In comparison with other mammalian cellular proteins, selenoproteins are unique in that they have a co-translational insertion of a rare amino acid called selenocysteine. Selenocysteine is similar to cysteine in the exception that instead of a sulfur component in the sidechain of the amino acid, there is a selenium ion [8]. The biosynthesis of selenocysteine starts with the charging of the tRNA for selenocysteine, tRNAsec. This tRNA gets charged initially with serine by the seryl-tRNA synthetase. This is not an efficient or perfect process, with this only occurring correctly in the human system about 10% of the time. The now charged seryl-tRNAsec can be converted to selenocysteyl-tRNAsec by using selenophosphate as a donor [9]. Bioavailable forms of selenium coming from the diet/environment like sodium selenite can also act as a donor for this process, hence correlating to an increase in selenoprotein expression in mammals with excess selenium in their diets [10]. Once all of this has occurred, the tRNAsec gets co-translationally inserted into a mRNA coding for a selenoprotein at a highly conserved UGA stop codon. This is able to occur because all selenoprotein mRNA has a region of

highly conserved nucleotides in their 3' Untranslated Regions (UTR) called a Selenocysteine Insertion Sequence Element (SECIS Element). The SECIS Element forms a conserved hairpin loop that recruits specialized binding protein SECIS Binding protein, which then recruits an elongation factor for selenocysteyl-tRNAsec. All of this allows the ribosome to readthrough the UGA stop codon during translation of a selenoprotein and treat it as the codon and insertion site for a selenocysteine amino acid (**Figure 1**) [4, 11, 12]. This is an incredibly special process, because not only do all of these factors have to come together, but the stop codon has to be the UGA stop codon. There are three stop codons in total including the UGA stop codon; UAA and UAG are not recognized as a potential selenocysteine insertion sites [11]. Due to all these factors, selenoprotein expression can be considered a complex, highly regulated and important process in the cell. Selenoproteins have been shown to be vital to cellular function. Cells that do not express selenoprotein Thioredoxin Reductase 1 (TXNRD1) undergo apoptosis [13]. Glutathione Peroxidase 1 (GPx-1) has been shown to be vital for immune response against viruses [14]. Selenoprotein expression increases when there are more Reactive Oxigenated Species (ROS) in the cellular environment in order to maintain homeostasis [14]. Selenoprotein P is the main source of selenium and selenium transport in mammalian systems, and several severe neurological conditions are related to or compounded by selenium deficiencies or total inability to express if fully [15-18]. So, we know that not only is the translation of selenoproteins complex involving multiple components that need to come together, but we also know that their expression is vital for cell and organism health. Due to the role selenoproteins play in viral response of the cell, it is not surprising that evidence surrounding viral interactions with selenoprotein mRNA's, and the effects viruses and selenoproteins have on each other in a host cell is prevalent in scientific literature [2, 14, 19-21].

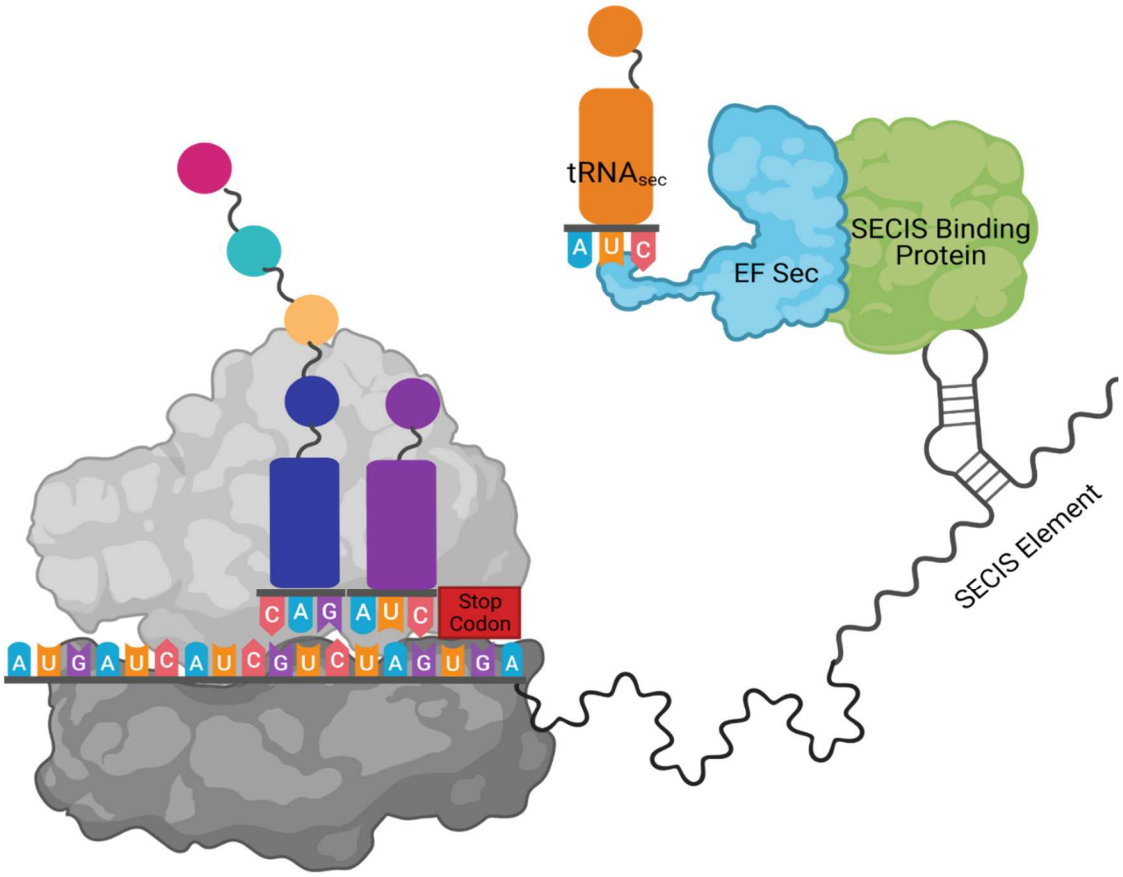


Figure 1: Post-translational insertion of a selenocysteine via SECIS Element recruitment of SECIS Binding Protein, EF Sec, and tRNA$_{sec}$. When an mRNA coding for the expression of a selenoprotein gets translated, the conserved SECIS Element in the 3' UTR region (black) forms a secondary structure that recruits binding proteins (green and light blue) that incorporate tRNA$_{sec}$ (orange) and allow the UGA stop codon (red) to be read through as a selenocysteine codon and continue on with protein translation versus stopping translation. 5' end of mRNA has been fed through the ribosome (gray), and random tRNA's (dark blue and purple) are in the P and A sites respectively in the ribosome during translation. Image is not to scale and was made using Biorender.

1.2 Viral Impacts

Viruses are a diverse group of microorganisms that have the potential to wreak havoc on a global scale and dramatically shift cultural norms and world economies. Viruses have played a key role in placing us where we currently reside in the world today and continue to impact where we are going to go and when. The examples of this are vast and too involved to fully delve into here, but just a few examples involve the Black Plaque wiping out a third of the European population [22].

The first English settlers unleashing smallpox on the New World, and in the process decimating Native American populations and paving the way for British colonies [23]. Haunting images of hospital rooms filled with iron lungs as many individuals battled the last stages of rampant Polio infection still are printed in our history books [24]. The 1918 Spanish Flu had an immense impact on civilization at a global scale. Some historians even speculate that it may have tipped the outcome of World War I, as it affected the armies of Germany earlier and more virulently than their Allied opponents [25]. Cultural scars resulting from the HIV/AIDS epidemic in the 1980s and the millions left dead due to ignorance and bigotry over the virus still remain in the LGBTQ+ community. The Ebola outbreak rampaging Africa with terrifying mortality rates and symptomology from 2014-2016 [26]. Zika infection continues to spread through the Americas, and the secondary complications associated with it are devastating and currently, the only way to avoid them is to avoid areas with known cases of Zika infection [27]. These are just a few of the more famous viral historical events that have shaped our current global positions. Viruses get passed to us in the air, through our food, through animals/insects and through our gatherings with each other. We cannot avoid them. Even without including the current COVID 19 pandemic, the global impact viruses have and how they have shaped modern day society and practices is undeniable; and this current pandemic has not only completely annihilated our sense of normalcy today, but an entire generation is growing up isolated/remote because of it. We have no way of understanding how that will affect the cultural landscape in the future as these pandemic children enter adulthood. Specifically, for this paper, we will be talking about 2 viruses, Zika Virus (ZIKV) and the Human Immunodeficiency Virus (HIV). What we are currently going through with the COVID 19 pandemic serves as a period to a long argument history has written us on how understanding viruses is vital to the continued survival of our economies, social gatherings,

and lives. We need to understand how viruses interact with humans on a molecular level to improve identification of viral infection, reduce the infection rate and severity of infection, and to prevent infection/transmission to protect future generations.

1.3 Viruses and their survival strategies

Viruses are among some of the smallest microorganisms characterized today. They consist of a relatively small genome encoding usually no more than 10-30 genes, although some viruses do have larger genomes. Viral genomes can either be DNA or RNA based, and these genomes can be protected by a collection of glycoproteins and fatty acids called a viral envelope, but this is not always the case. An envelope aids in the virus binding to and entering cellular membranes through membrane bound proteins and receptors [28]. For RNA based viruses, they can either be positive sense or negative sense oriented. Positive strand RNA based viruses can be translated into a long viral polypeptide shortly upon infection into the cell. Negative strand viruses have to be converted into DNA, then RNA using viral reverse transcriptase before they can be translated into viral proteins [29]. Due to the fact that viruses have such small genomes, they don't have the capacity to express proteins on their own that they would need like cytoskeletal proteins such as Beta-actin or their own ribosomes. Without the exploitation of pre-existing host cellular proteins and processes, viruses would not be able to propagate themselves and would cease to exist as we understand them today [28]. Viruses also employ various strategies within their nucleotide sequences, enabling them to extend their genetic code past what it would seem capable of expressing. There are 3 reading frames for genetic code, the -1, 0, and +1 reading frames. Viruses utilize a phenomenon called frameshifting during ribosomal translation so that they can change the reading frame the ribosome reads as it translates the viral RNA into a viral polypeptide chain. This results in the ability to express elongated or completely different proteins

[30, 31]. They are able to manage this a few different ways, one being conserved slippery sequences (tandem adenine or uracil nucleotides (UUUAAA) which do not interact as strongly with the ribosome) allowing the ribosome to "slide" into a +1 reading frame, and another being strongly conserved nucleotide regions in the 3' UTR of the viral RNA. These conserved regions form complex secondary structure that is vital to the virus in many ways, one being creating steric hinderance for the ribosome, which can push it into a -1 frameshift [32, 33]. **Figure 2** shows what this would look like. Some proteins in viruses aren't even able to be expressed without a frameshift, as is the case with one of the viruses discussed in this dissertation, HIV-1. Without a frameshift into the -1 reading frame, HIV-1 wouldn't have the capacity to express *pol* gene products, which are critical for the synthesis and integration of viral DNA into the host genome and the generation of capsid proteins [34]. With *flaviridae* family of viruses (of which the other virus of topic in this dissertation, Zika, is a member of), their conserved secondary structure in the 3' UTR in their genomes allows for host cell rnase Xrn1 to pause, which allows the non-degraded remains of their genome to break free in the form of subgenomic *flaviridae* RNA (sfRNA). This goes on to help the virus avoid host cell immune degradation by interfering with interferon generation and response. Antisense interactions between host cell mRNA and viral RNA are also a strategy employed by viruses to knockdown host cell protein expression that would aid in a viral immune response [35-38]. In this dissertation, evidence is provided that antisense interactions between viral RNA's and host cell mRNA's can also allow for stop codon UGA readthrough. There is mounting evidence to suggest that viruses employ these antisense interactions especially with regards to selenoprotein mRNA's in order to utilize host cell selenoprotein translational machinery to insert their own selenocysteines co-translationally into their viral proteins (**Figure 3**) [12, 14, 19, 39-41]. We suspect and will provide evidence that this

could be happening in ZIKV infected or HIV-1 infected cells. These strategies seem esoteric and random with minimal rates of success but seem to have huge effects on viral proliferation and virulence, hence why the nucleotide regions associated with these various viral hacks are so heavily conserved across viral strands.

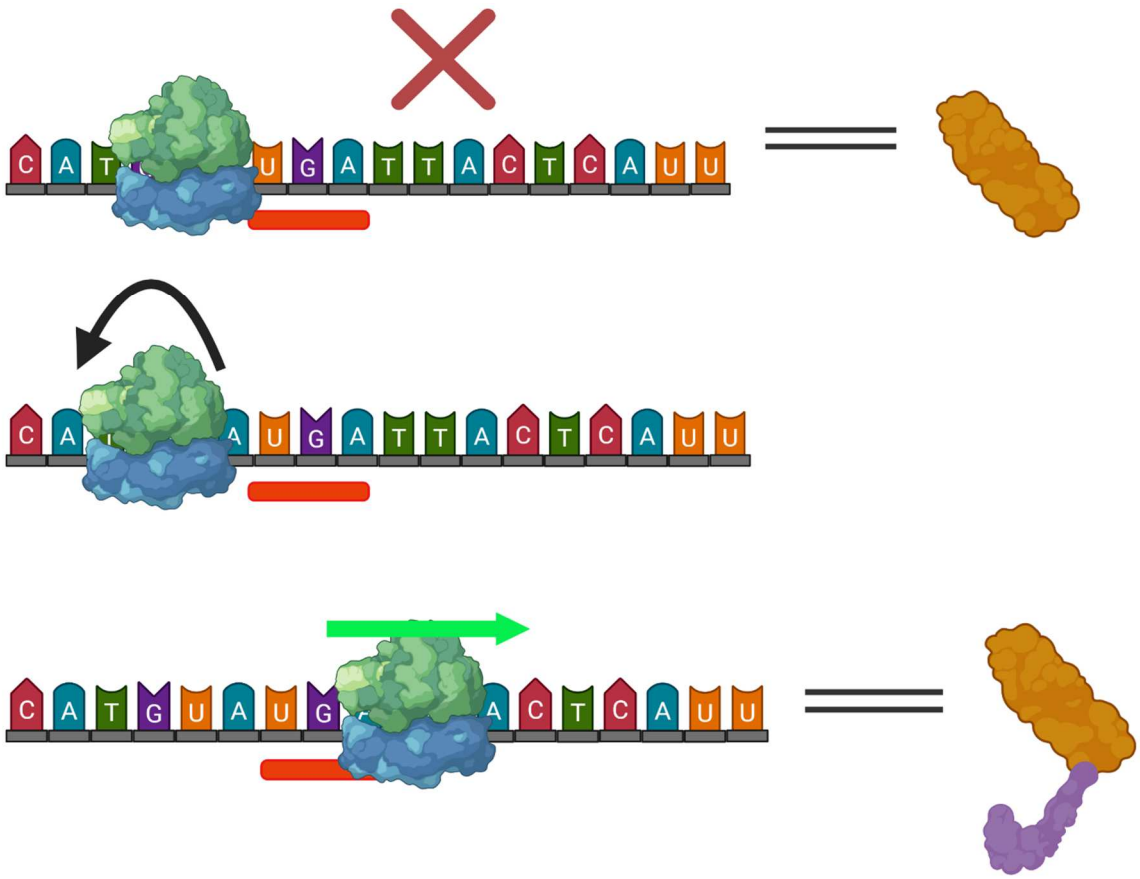


Figure 2: Viral strategies utilized to create extensions on translated proteins past UGA stop codons. Visualization of possible protein products that a stop codon readthrough could yield. The top strand of mRNA is being read by the ribosome in the 0 reading frame where the ribosome (blue/green) reads the stop codon (red) as a stop codon and translates the protein (yellow) coded by that mRNA. In the middle panel, the ribosome translates the same mRNA, only a frame-shifting event has happened that has caused the ribosome to slip back in the -1 reading frame (black arrow) so that the stop codon (red) will no longer read as a stop codon but as a codon for an amino acid. The bottom panel is the result of the new elongated protein (yellow/purple) that was able to be translated by the ribosome once it was frame-shifted out of the 0 reading frame which allowed it to readthrough the stop codon.

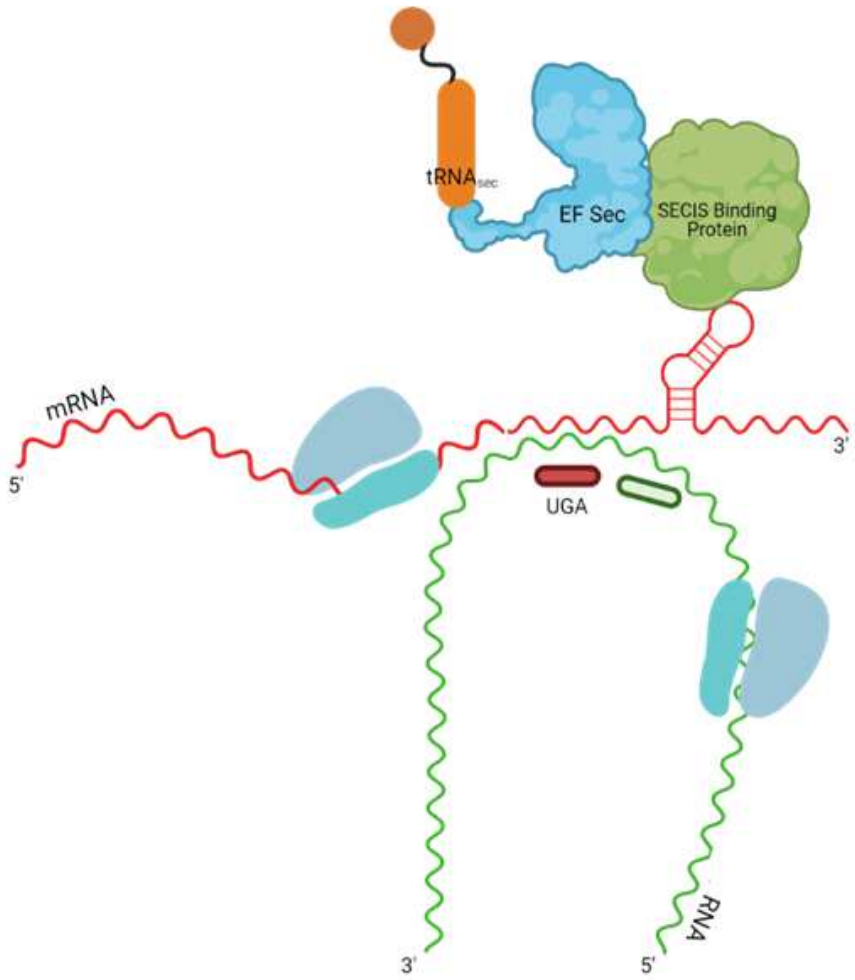


Figure 3: Visual into theorized utilization of host selenoprotein translation machinery to synthesize viral selenoprotein. The green RNA species is viral RNA and the red RNA species is a host cell selenoprotein mRNA. Visualized in red is a highly conserved UGA stop codon just after a viral gene and before the 3'UTR of the viral RNA. The hair pin loop at the 3' end of the selenoprotein mRNA represents a SECIS unit, which has already recruited SECIS Binding Protein, EF Sec, and $tRNA_{sec}$. Here the viral RNA is interacting in an antisense fashion in a highly selective manner in order to use the mammalian selenoprotein translational machinery to readthrough its own stop codon and extend the viral protein (green) into a seleno-isoform of that protein.

CHAPTER II

INHIBITION OF SELENOPROTEIN SYNTHESIS BY ZIKA VIRUS MAY CONTRIBUTE TO CONGENITAL ZIKA SYNDROME BY MIMICKING SEPP1 KNOCKOUT AND GENETIC DIEASE PCCA

Gabrielle P. Dailey[a], Ethan W. Taylor[a], Jan A. Ruzicka[b]

[a]Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, 435 Patricia A. Sullivan Science Building, PO Box 26170, Greensboro, NC 27402-6170, United States of America. [b]Department of Basic Pharmaceutical Sciences, Fred Wilson School of Pharmacy, High Point University, One University Parkway, High Point, NC 27268, United States of America.

2.1 Abstract

Selenium status plays a major role in health impacts of various RNA viruses. We recently reported potential antisense interactions between viral mRNAs and host mRNAs of the antioxidant selenoprotein thioredoxin reductase (TR). Here, we examine possible targeting of selenoprotein mRNAs by Zika virus (ZIKV) as a pathogenic mechanism, because microcephaly is a key manifestation of Progressive Cerebello-Cerebral Atrophy (PCCA), a genetic disease of impaired selenoprotein synthesis. Potential antisense matches between ZIKV and human selenoprotein mRNAs were initially identified via nucleotide BLAST searches, using ZIKV genomic RNA as a probe. The strongest antisense matches of ZIKV regions, against human TXNRD1 and selenoprotein P (SePP1), were validated by algorithms for prediction of RNA hybridization and microRNA/target duplexes. The ZIKV-SePP1 interaction was further assessed

Western blot probe for SePP1 and TXNRD1 expression as well as RT-qPCR analysis was conducted on non-ZIKV infected and ZIKV infected cells and effects of selenium supplementation on any protein or mRNA change in SePP1 and TXNRD1 expression upon ZIKV infection was observed. Computationally, ZIKV has regions of extensive (~30bp) and stable ($\Delta E < -50$kcal/mol) antisense interactions with mRNAs of both TXNRD1 and SePP1, a selenium carrier protein essential for delivery of selenium to the brain. The ZIKV/SePP1 hybridization was experimentally confirmed at the DNA level using synthetic oligonucleotides to ensure that the proposed interaction between ZIKV RNA and SePP1 mRNA was possible *in vitro*. ZIKV infection was shown to knock down SePP1 expression by 99.1% and TXNRD1 expression by 90.4% in ZIKV infected cells. Selenium supplementation seemed to have little effect on any TXNRD1 or SePP1 protein expression in the presence of ZIKV infection. SePP1 mRNA increases by 6 folds upon infection of ZIKV, and TXNRD1 mRNA increases nearly 8 fold with ZIKV infection and selenium supplementation. These findings provide more insight into how ZIKV infection can lead to microcephaly onset and how viral RNA and host cell RNA interactions serve as a key strategy for viral inhibition of host cell viral immune responses.

2.2 Introduction

Zika virus (ZIKV) is an enveloped, positive sense, RNA based virus, and was first isolated near the Zika forest by Lake Victoria in Uganda, Africa.

ZIKV is a member of the *flaviridae* family of viruses; the viruses associated with Dengue fever, yellow fever, and Japanese encephalitis are also *flaviridae* family members. ZIKV is an arborvirus meaning a large part of its transmission route is through the bite of an *aedes* mosquito (*A. aegygpti, A. albopictus, A. africanus,* and *A. luteocephalus*), but it can also be transmitted sexually, through saliva, through blood transfusion, and in a pre-natal fashion from mother to

fetus. Symptoms of ZIKV infection tend to be relatively mild for the majority of healthy, non-immunocompromised adults; but can include fever, fatigue, myalgia, headaches, rashes, conjunctivitis and retro-orbital pain [42, 43]. Where ZIKV becomes a larger concern is in regard to the secondary conditions that have been associated with infection including Guillain-Barre syndrome in adults and ZIKV congenital microcephaly in infants whose mothers contracted the virus during pregnancy.

Guilian-Barre syndrome is an autoimmune condition that is characterized by the afflicted individuals immune system attacking the myelin sheath surrounding the axons to their neuronal cells, leading to symptomology including limb paralysis, bilateral facial palsy and autonomic dysfunction with sustained ventricular arrythmia, hypotension, seizures, general motor-cortex degeneration, and at times death [42].

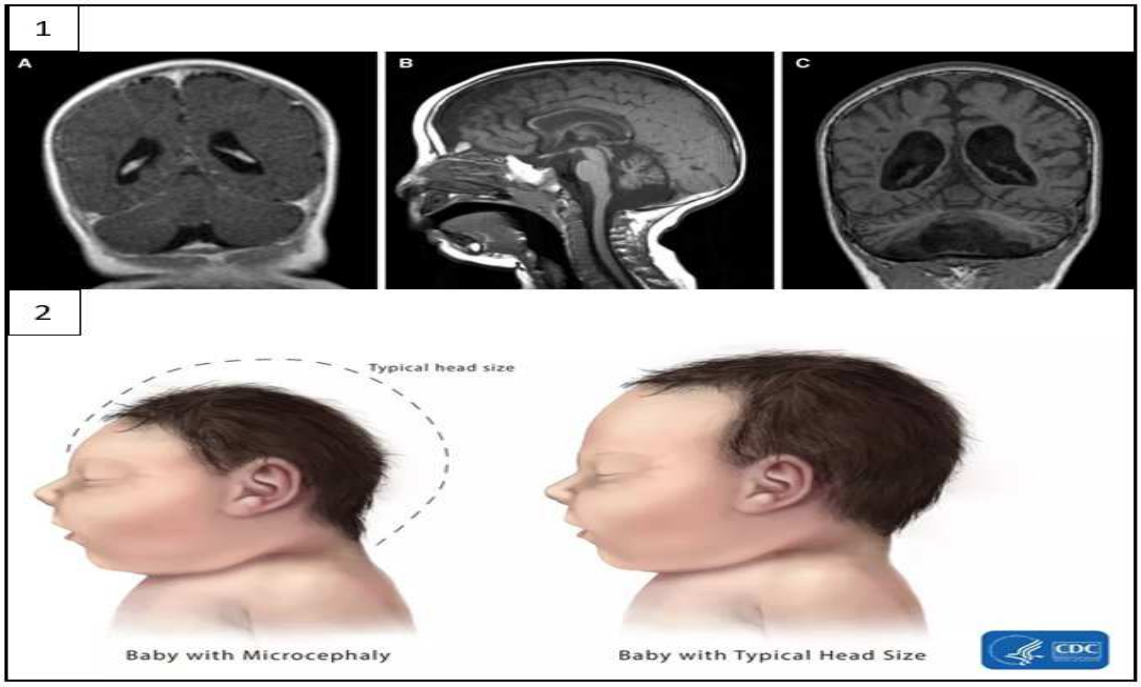


Figure 4: Comparison of phenotypes of PCCA and ZIKV congenital microcephaly. Panel 1 shows MRI of child with PCCA and demonstrates reduced cranial circumference and compression to the brain along with abnormal development [1]. Similar in phenotype and symptomology, Panel 2 shows phenotype of ZIKV congenital microcephaly as defined by the CDC.

This condition is not always reversible, and the afflicted individual could be affected for the duration of their lives. ZIKV congenital microcephaly is the focus of this chapter. Microcephaly is defined as a clinical condition where the infant head circumference is significantly smaller than what is normal and is considered a severe congenital defect. Two generalized forms of microcephaly are recognized, one being diagnosed around 32 weeks of the gestation period and characterized by a decrease in neuron production, and the other characterized by a lack of head growth after birth due to loss of dendritic connections [43-47]. A 2012 study also linked the depletion of radial glial cells and neuronal stem cells (neuronal progenitor cells) via cell death or premature differentiation to microcephaly onset in utero [48]. A rare genetic disease that causes microcephaly is Progressive Cerebello-Cerebral Atrophy (PCCA). It is an autosomal recessive syndrome characterized by profound mental retardation and severe quadriplegic spasticity, and its cause is a defect in selenium metabolism. Specifically, the mutations responsible for this disease were found to be in the SepSecS gene, which encodes O-phosphoseryl-tRNA:selenocysteinyl-tRNA synthase, a selenium transferase that catalyzes the final step in the biosynthesis of Sec. Unique among the amino acids, Sec is synthesized while bound to its cognate tRNA, by conversion of O-phosphoseryl-tRNASec to selenocysteinyl-tRNASec. Homozygous individuals, lacking a functional copy of SepSecS, have therefore lost the unique catalytic benefits of Sec in selenoenzymes, at best having a serine hydroxyl group in place of the selenol of an active site Sec, leading to a drastic or total loss of catalytic activity in thioredoxin reductases, Se-dependent glutathione peroxidases, iodothyronine deiodinases, and other human selenoproteins. [1, 44]. This means that they lack the ability to synthesize any selenoproteins, which are a class of proteins fundamentally important for immune response, redox reactions, and neuronal development and differentiation within the body [49, 50].

Zika related microcephaly onset is most critical during the first trimester but usually isn't diagnosed until the second trimester and occurs when a pregnant individual becomes infected with ZIKV and vertically transmits it to the fetus in utero [42]. Symptoms include profound mental retardation and severe quadriplegic spasticity like PCCA. The exact mechanism of vertical transfer from mother to fetus is still not fully characterized, but ZIKV has both been shown to infect/be isolated from placental cells, amniotic fluid and fetal brain cells including multipotential neuronal progenitor cells and glial cells [51-54]. Zika related microcephaly is sometimes not diagnosed until after birth, which could be due to more stringent screenings and diagnostic measures being needed, or post-utero ZIKV infection perinatally through breast feeding [55, 56]. There is not a clear answer currently in the literature as to why ZIKV infection leads to the onset of microcephaly in some ZIKV+neonates, but it is the goal of this chapter to address some gap in the knowledge surrounding this phenomenon.

In this chapter, a case is made linking the symptomology and cause of PCCA to ZIKV congenital microcephaly. PCCA and ZIKV congenital microcephaly have remarkably similar phenotypes and consequences for the afflicted individuals (**Figure 4**). As stated, PCCA is caused by a complete lack of ability to express any selenoproteins; and through bioinformatic methods, our group identified two predicted antisense regions of interaction between ZIKV RNA and host cell selenoprotein mRNA coding for the expression of both Selenoprotein P (SePP1) and Thioredoxin Reductase 1 (TXNRD1) (**Figure 5**). SePP1 is predominately secreted from the liver but is expressed in many tissues throughout the body, and it is the sole source of selenium transport throughout the body [6, 57, 58]. SePP1 has 10 selenocysteine residues and transports selenium to the brain by binding the Apo2ER receptor, where it will then get broken down in the cell by protein selenium lyase; selenium lyase being essential for the recycling of selenocysteine

in the body for the continued synthesis of $tRNA_{sec}$ [15, 59]. SePP1 protein expression knockdown has been linked to several neurological conditions in mouse models including neuron degeneration, brain stem deterioration, seizures, and death [59]. Neuronal cells express a truncated isoform on SePP1 on their own, but studies have shown that mice only expressing the truncated isoform of SePP1 in their neuronal cells and not expressing full length SePP1 from their livers are more dependent on selenium in their diets and are not able to circumvent the negative effects of SePP1 knockout on a selenium deficient diet [60, 61]. TXNRD1 is part of the enzymatic chain that converts RNA sugars to DNA sugars. It is essential for cell homeostasis, and when TXNRD1 knockdown is seen in cell lines, apoptosis is induced. TXNRD1 is expressed 7 times more in fetal multipotential neuronal progenitor cells than it is in adults and is also essential for cell differentiation [62, 63].

Like the majority of other viruses, ZIKV has a relatively small genome (10,794 ribonucleotide bases) encoding for only 10 genes (3 structural and 7 non-structural proteins). As discussed, viruses often employee various methods to get the biggest translational/expression payload from their relatively modest genomic bank including hijacking host cell machinery or by employing different reading frames, secondary RNA structures, frame shifts or alternative splicing [2]. Antisense interactions between viral RNA and host cell mRNAs are a characterized tactic that viruses employee to further interfere with host cell protein expression. Our group has identified two antisense regions of predicted high affinity interaction between SePP1 mRNA (the predicted interaction on the exon can be extended 30 nucleotides in the 3' direction to span an intron/exon junction in SePP1 pre-mRNA as well seen in **Figure 5**) and TXNRD1 pre-mRNA (spanning an intron only seen in **Figure 5**). Since members of the *flaviridae* family of viruses can replicate in the nucleus, it is possible that these predicted antisense interactions are occurring in the nucleus,

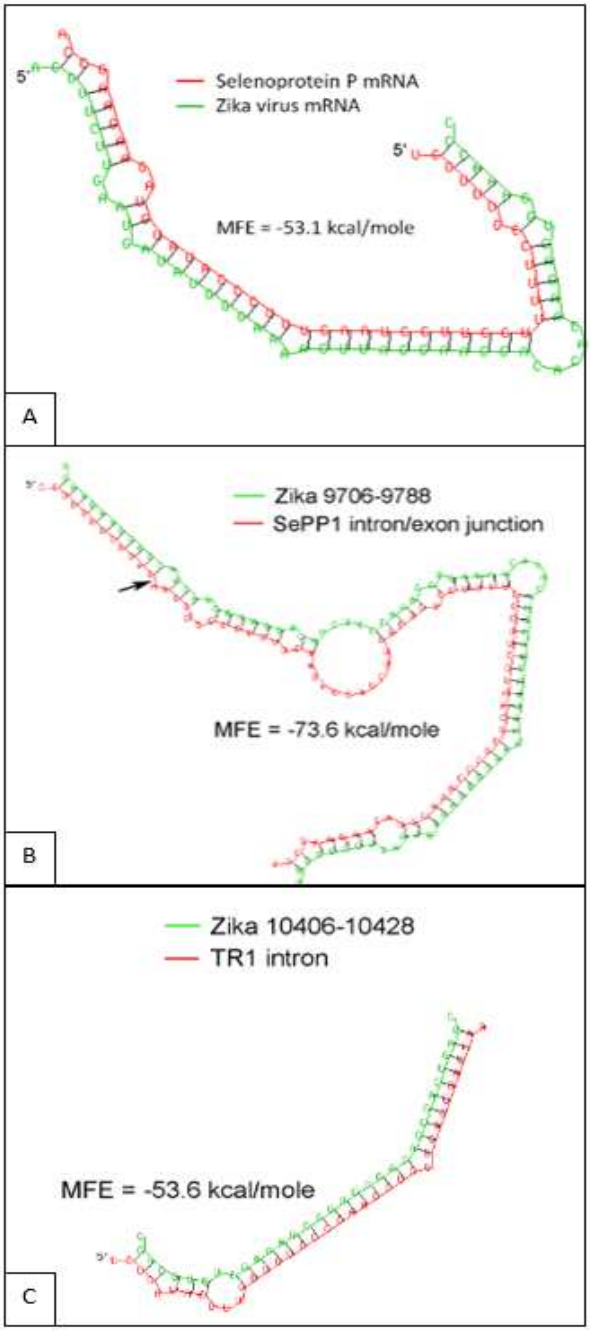
knocking down protein expression of TXNRD1 and SePP1 [54, 64]. The proposed antisense interactions are occurring between these host cell RNA species and the 3' Untranslated Region (UTR) of ZIKV RNA. *Flaviridae* viruses have conserved nucleotide sequences in their 3' UTR that form highly conserved pseudoknot secondary structures. They employ this tactic to slow down host cell degeneration of the foreign viral RNA by host cell rnase Xrn1. The secondary structure in the 3' UTR of the viral RNA causes the rnase to pause, releasing the remaining non-degraded viral RNA into the cytosol. This subgenomic *flaviviral* RNA (sfRNA) interferes with host cell interferon production and increases viral propagation by lowering cell viral defense mechanisms [35, 36, 38, 65-67]. Hence, the proposed antisense interactions could be favorable for ZIKV proliferation in the cell. Here, a case is made that predicted antisense interactions (**Figure 5**) between ZIKV RNA and SePP1 mRNA and TXNRD1 pre-mRNA are occurring in ZIKV infected cells, and that this interaction is interfering with the already complicated selenoprotein expression complex, thus knocking down expression of these two proteins. Due to their vital function for neuronal development and maintenance to the fetus, the knockdown of these proteins could be compounding factors in the onset of ZIKV congenital microcephaly in ZIKV+neonates.

Figure 5: Predicted antisense interactions between Zika virus mRNA and regions either in introns or spanning RNA slice sites of human selenoprotein mRNAs

A: The hybrid dsRNA secondary structure shown, and computed interaction energy were generated using the RNAHybrid 2.2 program. When used as a query vs. the entire SePP1 mRNA, the illustrated (green) 54 base region of Brazilian ZIKV strain SPH2015 (bases 9706-9759) yields the match shown as the minimum free energy (MFE) antisense interaction, to a (red) 47-base region of SePP1 (bases 541-587; these sequence ranges include the unpaired single base overhangs shown at each end).

B: Spanning an intron/exon junction of SePP1. Flaviviruses like ZIKV replicate in the nucleus as well as in the cytosol, meaning ZIKV mRNA could also have the opportunity to interact with unspliced cellular pre-mRNAs. The ZIKV/SePP1 interaction shown Panel A can be extended by another ~30 bases at the 3' end of the ZIKV sequence, into an intron of SePP1, which is the interaction shown in this panel. The black arrow indicates the intron/exon boundary in SePP1. The right side of the structure pictured here is identical to the structure shown in Panel A, only rotated counterclockwise. Formation of such a structure might inhibit the splicing and maturation of SePP1 mRNA. An essentially identical antisense complex is predicted using the IntaRNA progam, differing only in that the terminal 7 base pairs at the bottom of the structure shown above are not included in the IntaRNA prediction (**Appendix A**, **Figure 22**).

C: Targeting an intron of TR1: Both the RNAHybrid and IntaRNA programs predict the antisense interaction shown here between a 3' noncoding region of ZIKV mRNA (near the beginning of the well documented non-coding "subgenomic Flavivirus RNA" region) and an intron of TR1. An identical antisense complex is predicted using the IntaRNA progam.



2.3 Materials and methods

2.3a Computational methods for identification of virus-host RNA antisense interactions

Potential antisense matches between regions of ZIKV and host selenoprotein mRNAs were initially identified via nucleotide BLAST searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi). A complete 2015 Brazilian ZIKV genomic mRNA sequence (see following section) was used as a probe for Nucleotide BLAST (blastn option) with default search parameters, initially against the Reference RNA sequence database (refseq.rna), restricted to Homo sapiens (taxid:9606) with selenoprotein as a search term. Because ZIKV is known to replicate in the nucleus, the search was later extended to include genomic sequences (corresponding to unspliced RNA; see below). The initial mRNA search led to the identification of the core region of the antisense match to a coding region of SePP1 mRNA shown in **Figure 5**. The RNAHybrid program (http://bibiserv2.cebitec.uni-bielefeld.de/rnahybrid) was then used to assess the match using an algorithm and parameters designed for actual RNA:RNA interactions, and to investigate the possibility that, with some mismatches and bulges characteristic of dsRNA, the match at the RNA level might be extended in the 5' and 3′ directions beyond the core region identified by BLAST [68]. A more stringent method for accurate prediction of RNA-RNA interactions was then applied: the IntaRNA program (http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp) [68-70]. This algorithm factors in not only the hybridization energy of the interacting pair of RNAs, but also takes into consideration the "unfolding energy" required to overcome and outweigh the stability of any internal RNA secondary structures within the two individual RNA strands. The program will only identify a match if the net inter-strand binding energy is substantially lower (i.e., more stable) than the sum of the internal folding energies of the individual RNA strands. Using input fragments of up to 2000 bases in length (the maximum

supported by the program) IntaRNA was in each case able to identify as the single most significant match an antisense interaction that was essentially identical to that found by the combined BLAST-RNAHybrid approach.

2.3b Reference sequences used in antisense sequence analysis and oligonucleotide design

The ZIKV Brazilian strain SPH2015, Genbank accession number KU321639, was used as a reference sequence for ZIKV, as it was the only complete Brazilian genomic sequence available at the time the study was initiated. Using methods described below, the most significant antisense match found to a region of a human selenoprotein was to selenoprotein P (SePP1). The initial SePP1 hit was to a region of mRNA common to all of its transcript variants; the sequence numbering used in the figures and text is from transcript variant 6 (the longest variant), Genbank # NM_005410.2. Because the antisense match to the SePP1 mRNA was at the very beginning of an exon, to explore the possibility that this match might extend into the upstream intron in unspliced RNA, that search was repeated using genomic DNA instead of the mRNA database, leading to the identification of an extended match spanning the intron-exon boundary in the SePP1 genomic sequence, Genbank DQ022288.1. An antisense match of a 3' noncoding region of ZIKV to an intron of thioredoxin reductase 1 (TXNRD1) was also identified in the search vs. human genomic DNA; the TXNRD1 sequence numbering used in the figures and text is from the complete TXNRD1 gene sequence, Genbank # DQ157758.1.

2.3c Experimental confirmation of the predicted ZIKV-SePP1 antisense interaction at the DNA level

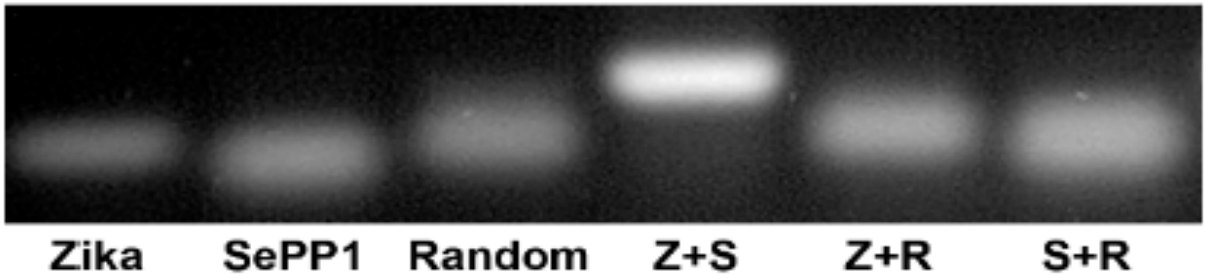


Figure 6: Target-specific in vitro DNA hybridization of the predicted ZIKV-SePP1 antisense pairing. Confirmed at the DNA level by gel shift assay, using DNA oligonucleotides corresponding exactly to the sequences **Figure 5**. The left three lanes contain only a single (unpaired) ssDNA oligo, as follows: Zika=the 54 base fragment from **Figure 5**, SePP1=the 47 base fragment from **Figure 5**, Random=a 47 base randomly shuffled version of the SePP1 fragment. The right 3 lanes are from incubations of pairwise combinations of those 3 oligos: Zika+SePP1 (Z+S), Zika+Random (Z+R) and SePP1+Random (S+R). Of these, only the Zika+unshuffled SePP1 hybridize to form a slower moving dsDNA band that migrates as expected for ~50bp dsDNA (size markers not shown; for dsDNA ladder, see original uncropped gel photo, Fig.S5 in online Supplemental Materials). The Z+R and S+R combinations still migrate as single stranded DNA, demonstrating the specificity of the Z+S antisense interaction.

An electrophoretic mobility shift assay was used to confirm the predicted interaction shown in **Figure 5** via demonstration of in vitro DNA hybridization of the cognate ZIKV:SePP1 pair of sequences. The procedure was essentially as described previously [2]. Briefly, synthetic single stranded ssDNA oligomers (Integrated DNA Technologies, Inc., Coralville, IA) were obtained, corresponding precisely to the ZIKV and SePP1 sequence fragments shown in **Figure 5**. An additional oligo with a random sequence of identical base composition to the SePP1 fragment was used as a control. Prior to gel electrophoresis, oligos (~1 μg each in 10 μl PBS), either singly or in all three possible pairs, were incubated at 37°C for 15 hours, followed by cooling to room temperature over 1 hour. See legend to Fig. 1 for details about the arrangement of lanes on the gel. Bands were separated on a 4% agarose gel and visualized using ethidium bromide (**Figure 6**).

2.3d Cell Culture

Cell culture media was prepared by combining 444.5 mL of 1x DMEM media (Gibco™ catalog number 11054020), 50 mL heat-inactivated FBS (Gibco™ catalog number 10082147), 5 mL 100X L-glutamate (Invitrogen catalog number 25030081), 500 uL Penicillin-Streptomycin (Gibco™ catalog number 15070063) and filtering it through a 0.22 micron combined filter/container apparatus. HEK 293T cells were defrosted from original ATCC stock and resuspended in 5 mL of cell culture media in a T25 flask (vented breakneck flask) and allowed to come to confluency over the course of 2 weeks at 37^0C and 5% CO_2. 1X PBS washes and media changes were performed during this growth cycle every 3 days. Once cells reached 70-90% confluency, the cells were passaged by incubating 1 mL of TrypLE™ Express Enzyme (1X, no phenol red, Gibco™ catalog number 12-604-013) on the cell monolayer at 37^0C and 5% CO_2 for 5 minutes. 1 mL of cell culture media was added to deactivate the TrypLE™ Express Enzyme and the cell mixture was centrifuged at 1,800 rcf for 4 minutes at room temperature. The resulting cell pellet was resuspended in 1 mL of cell culture media. 10 uL was removed from the new cell mixture and added to 10 uL of 0.4% Trypan blue solution (Gibco™ catalog number 15250061). 10 uL of the resulting solution was analyzed on the Countess™ II (Invitrogen™ catalog number AMQAX1000) for cell viability and cell count calculations. 4 aliquots of 1 x 10^5 cells were removed from the cell mixture and seeded into 4 separate T25 flasks and allowed to grow under the same conditions previously specified until the cell monolayer reached 20-30% confluency. Once desired confluency was reached, the 4 flasks were separated based on experimental conditions HEK alone (positive control), HEK with 20 nM sodium selenite supplementation, HEK with ZIKV infection, and HEK with ZIKV infection and 20 nM sodium selenite supplementation. The two HEK experimental conditions not destined for ZIKV infection

underwent a PBS wash and 1 mL of Opti-MEM™ Reduced Serum Media (Gibco™ catalog number 31985070) was added to each flask and allowed to sit overnight in the 37^0C and 5% CO_2. The other two experimental conditions that called for ZIKV infection were washed with 1X PBS and 1 mL of ZIKV infection inoculum (MOI 0.01) was added to the cell monolayer and allowed to sit overnight at 37^0C and 5% CO_2. After the overnight period, the Opti-MeM and ZIKV infection inoculum from their respective experimental conditions were decanted into 10% bleach. All experimental conditions underwent a 5 mL PBS wash. Experimental conditions calling for sodium selenite supplementation received 5 mL of cell culture media with a 20 nM concentration of sodium selenite and experimental conditions not calling for sodium selenite received 5 mL of cell culture media only. All experimental conditions were allowed to grow an additional 96 hours undisturbed at 37^0C and 5% CO_2. After 96 hours, the cell monolayers from each experimental condition were harvested following the same procedure described earlier with the cellular passage. Cell counts were performed on all experimental conditions, and the resulting cell mixtures were again spun down at 1,800 rcf for 4 minutes to form a cell pellet, where they were stored at -80^0C until further processing. All cell culture was performed in a Labgard ES Class II, Type A2 biological safety cabinet. All cell culture surfaces were exposed to UV light for 15 minutes before and after work was performed. All cell culture surfaces, disposables, and containers were wiped down with 70% ethanol before work and where applicable, after work was done. All cell culture involving live ZIKV infection was performed with BSL2+ PPE including gloves, lab coat, and eye protection. All cell culture was exposed to 10% bleach solution before being autoclaved at 14 PSI, 121^0C for 1 hour or until autoclave control tests turned positive for disinfection.

2.3e ZIKV Infection/propagation

PRVABC59 Zika viral stock was defrosted from original ATCC stock (VR-1843) and aliquoted into 10 100uL tubes and stored at -80^{0}C for future use. 100 uL of viral stock was added to 900 uL of Opti-MeM to achieve an MOI of 0.01. The resulting inoculum was placed directed onto HEK 293T cell monolayer cultured as previously described. The flask was gently rocked for 5 minutes before being placed in a separate CO_2 incubator overnight at 37^{0}C and 5% CO_2. The inoculum was then decanted and the cell monolayer was washed with 1X sterile PBS and 5 mL of cell culture media was added. The infected cells were allowed to culture undisturbed for a period of 96 hours, after which the cellular supernatant was collected and filtered through a 0.22 uM filter and stored in 1 mL aliquots at -80^{0}C for future use.

2.3f RT-qPCR

Table 1: Primers used for amplification of RNA targets

RNA amplified	Forward Primer in Primer Mix	Reverse Primer in Primer Mix
SePP1 mRNA	5'TCC TCC ATT CTA AAC TGC TAA TTA TCC3'	5'CCA GTG TTC TAT TTG CTT TAA TGA G3'
TXNRD1 mRNA	5' ACT CCG GCA TTT GCA GCA GAG3'	5' ATA AGG TCA TAG TCA TAG GAC TTG GG3'
Beta Actin mRNA	5'CAC CAT TGG CAA TGA GCG GTT C3'	5'AGG TCT TTG CGG ATG TCC ACG T3'
ZIKV RNA	5'AGA TGG AGC TTG TTG AAG TGG3'	5'CAT TCG TTT GAG CCT ATC CCA3'

Quick-RNA™ Miniprep kit (Zymo Research catalog number R1054) was used to extract and purify total RNA from cell pellets resulting from each experimental condition according to manufacturer's guidelines. 3 uL of the purified RNA was removed from each sample to be tested for concentration, integrity, and purity via nanodrop at A_{254}. 1 uL of SUPERase.In™ RNase Inhibitor (Invitrogen™ catalog number AM2694) was added to each RNA sample. Samples were stored at -20^{0}C. 500 ng of RNA from each sample were pipetted into a 48 well plate in quadruplicate for each experimental condition (HEK lysate, HEK lysate + Se supplementation, ZIKV infected HEK lysate, and ZIKV infected HEK lysate + Se supplementation). This was repeated 3 times to allow for amplification of all RNA targets. 2 uL of 0.2 uM primer mix (**Table**

1) was added to each well with 1 uL of SuperScript™ IV Reverse Transcriptase (Invitrogen™ catalog number 18090010), and 10 uL of PowerUp™ SYBR™ Green Master Mix (Applied Biosystems™ catalog number A25741). The reaction volume was brought to 20 uL with UltraPure™ Distilled Water (Invitrogen™ catalog number 10977-015). The well plate was sealed and cycled at 48^0C for 30 minutes for cDNA synthesis, 95^0C for 10 minutes for *Taq* Polymerase activication, and then cycled through 40 cycles of denaturing steps at 95^0C for 10 seconds and annealing/amplification steps 54^0C for 30 seconds. All one-step RT-qPCR data was gathered and analyzed using Eco™ Real-Time PCR machine and software (Illumina catalog number EC_900_1001). All RNA targets (SePP1, TXNRD1, and ZIKV RNA species) were run in quadruplicate for each experimental condition. The quadruplicate Ct value averages and standard deviations were calculated in Excel. P-values were determined by comparing experimental values (HEK+Se, HEK+ZIKV, HEK+ZIKV+Se) to the control HEK in a two-sampled, 2-tailed t-Test assuming equal variance in Excel.

2.3g Western Blots

1 mL of cold RIPA lysis buffer (10X Millipore, catalog number 20-188, diluted to 1X with deionized water) was added for every $5x10^6$ cells in the resulting cell pellets from each experimental condition. 10 uL of beta-mercaptoethanol, Halt™ Protease and Phosphatase Inhibitor Cocktail (100X, EDTA-free, Thermo Scientific™ catalog number 78441), and a 0.5 M EDTA solution (0.5 M EDTA in sterilized deionized water) was added to every 1 mL of cold RIPA lysis buffer prior to addition to the cell pellet. The cell pellet was resuspended in this solution and vortexed for 10 second intervals, three times each with 30 second intervals in between where the samples were incubated on ice. The samples were then sonicated for 15 second intervals, three times each with 30 second intervals in between each sonication where the

samples were incubated on ice. The sonicator probe was cleaned using 99% ethanol in between each sample to ensure no contamination between the experimental conditions. The samples were then placed on a 95^0C heating block for 5 minutes before being spun down at 16.1xg at room temperature for 10 minutes. The supernatant was collected and stored at -80^0C until further use.

The Pierce BCA Protein Assay Kit (Thermo Scientific, catalog number 23225) was used to determine relative protein concentrations of each experimental sample. The assay was done to manufacturers specifications and visualized at 560 nm on a plate reader. Equal amounts of protein from each experimental condition were loaded into a 12% Mini-PROTEAN TGX SDS-PAGE Gel (samples contained 1X PBS, protein sample, and 3 uL 10X Loading Dye (Innovating Science, catalog number IS5208)). 3 uL molecular weight ladder (SMOBio Enhanced 3-color*ExcelBand*™, catalog number PM2510) was added to one lane as a standard. The gel was set up in a BioRad Mini-PROTEAN vertical gel electrophoresis box with running buffer (10X Tris/glycine buffer, BioRad catalog number 1610732, diluted to 1X with deionized water) was added to the fill lines. An ice pack was placed in the gel box system and the setup was run at 80 V for 10 minutes and 115 V for 50 minutes. After running, the gel was removed from its cast and placed in deionized water along with 4 pieces of filter paper and 1 piece of 0.45 uM pore nitrocellulose membrane (BioRad catalog number 1620115) cut to the size of the gel. The wet sandwich was placed in the BioRad wet transfer casket and filled to the top of the set up with transfer buffer (1X running buffer, 20% methanol). An ice pack was placed inside the apparatus, and the transfer was allowed to run at 200mA for 1 hour and 15 minutes. Once the transfer was complete, the membrane was removed from the wet transfer casket and rinsed off once in deionized water before being blocked in 2% milk protein solution (2% w/v dried milk with 1X sterilized PBS) on a shaking table for 1 hour. The blocking solution was removed, and the blot

was washed 3X for 5 minutes each with 1X sterilized PBS before addition of primary antibody (either monoclonal anti-SePP1 in rabbit (Sigma-Aldrich catalog number SAB2103123-100UL), monoclonal anti-TT1 in rabbit (BETHYL catalog number A304-791A), or polyclonal anti-beta Actin (proteintech™ catalog number 20536-1-AP) depending on what protein expression was being measured). The blot was allowed to incubate with 10mL of 1:1000 dilution (in 1X sterilized PBS) of the primary antibody on a shake table at 4^0C overnight. Following this, the primary antibody was removed and the blot was washed an additional 3X with 1X sterilized PBS for 5 minutes a piece before 10mL of 1:5000 dilution (in 1X sterilized PBS) of secondary antibody (goat anti-rabbit IgG (H+L) Cross-Absorbed Secondary Antibody Alexa Fluor™ 700, Invitrogen catalog number A-12038 for SePP1 and TXNRD1 monoclonal antibodies and goat anti-mouse IgG (H+L) Cross-Absorbed Secondary Antibody Alexa Fluor™ 700, Invitrogen catalog number A21036 for polyclonal beta-actin antibody) was incubated on the blot for 90 minutes on a shake table at room temperature. The secondary antibody was removed and the blot was washed as previously described before being placed in sterilized 1X PBS and visualized on gel imager Odyssey CLx (LI-COR) excited at 700 nm with emission measured at 720 nm at the highest quality setting. The densitometry of the tagged protein bars on the resulting blots were analyzed using the gel reader software. Average protein expression for each target protein in each experimental condition was calculated by averaging the densitometry results of triplicate blots. Error bars were generated using standard error of the means and p-values were determined by comparing experimental values (HEK+Se, HEK+ZIKV, HEK+ZIKV+Se) to the control HEK in a two-sampled, 2-tailed t-Test assuming unequal variance. All calculations were performed in Excel.

2.4 Results and discussion

2.4a Western blot data

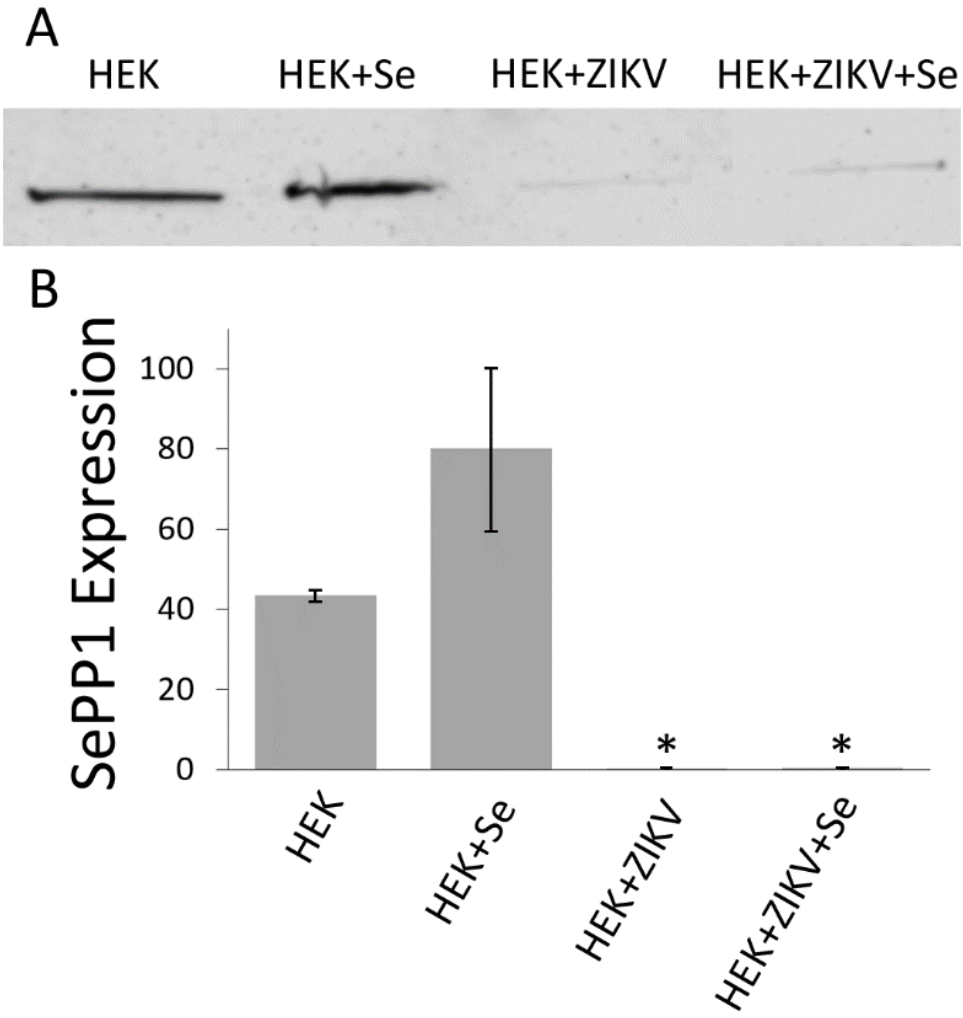


Figure 7: Western blot analysis of SePP1 Expression in ZIKV infected HEK 293T cells. A. *Western blot of SePP1 protein expression under several experimental conditions* (with or without addition of 20 nM sodium selenite, either $ZIKV^{+}$ or $ZIKV^{-}$). Equal concentrations of protein were used per sample. Lanes are designated in the images. Results show: **A.** Anti-SePP1-labeled protein bands across experimental conditions at the 60 kDa level, which corresponds to the predicted mass of full length SePP1 (isoform 1). Near-abrogation of SePP1 protein expression is seen in ZIKV-infected cells. **B.** *Quantitative expression of SePP1 in experimental conditions* (Y axis shows fluorescence intensity units/1000). HEK SePP1 expression starts at 43,500, increases to 80,100 for HEK+Se, drops to 268 for HEK+ZIKV, and increases slightly to 416 for HEK+ZIKV+Se. Error bars are ± standard error of the mean, * indicates a p-value ≤0.05.

Western blot analysis of SePP1 expression in HEK cells (**Figure 7, Panel A**) show expression of SePP1 across all experimental conditions at the 60 kdal line. The difference in the densitometry data for the SePP1 expression of the different experimental conditions was converted as a percent change from the baseline SePP1 expression calculated from the averaged results of the positive control HEK lysate only condition (**Table A1**, **Appendix A**). The results presented in **Figure 7**, **Panel B** and in **Table A1** were gathered by averaging the densitometry data from 3 western blots running each experimental condition. A dramatic decrease in SePP1 expression is observed in both experimental conditions where HEK cells were infected with ZIKV, and densitometry analysis of the SePP1 western blot shows a 99.4% decrease of SePP1 expression in ZIKV infected HEK cells with SePP1 only expressing at 0.6% in ZIKV infected cells compared to non-infected ZIKV infected cells (**Figure 7, Panel B**). This is essentially a complete knockdown of SePP1 expression in ZIKV infected cells. There have been 4 isoforms of SePP1 isolated from rat plasma [71-73]. All isoforms contain the first UGA selenocysteine insertion site in the N-terminal region of the protein, but have truncated C-terminal domains resulting with termination at positions of the 2^{nd}, 3^{rd}, and 7^{th} UGA insertion sites, resulting in fewer selenocysteine being present in these isoforms and being of a lower molecular weight [58]. SePP1a is the full length selenoprotein and has 9 selenocysteine residues in the C-terminal domain, resulting in a molecular weight of 60 kdal when glycosylated and run on an SDS-PAGE gel, thus indicating that the results observed are of the expression of full-length SePP1 [6, 57, 58]. Several studies have determined that selenoprotein P has a high turnover rate in not only rat plasma, but human plasma as well. This is in order to recycle the selenocysteine for the incorporation into other selenoproteins being translated in the cell [74, 75]. The efficient recycling mechanism of selenocysteine by the cell could be why no smaller isoforms of SePP1

were observed in any meaningful way in the ZIKV infected experimental conditions; as the cell may have been recycling selenocysteine by breaking down less effective SePP1 isoforms and using it for the translation of more relevant selenoproteins to the cell including TXNRD1. SePP1 is the main source of plasma selenium, and it has important functions for many tissues of the body [4-8]. However, studies have found that in SePP1 knockdown rats, SePP1 with isotopically labeled sodium selenite in the selenocysteine sites is most rapidly transported from the SePP1 knockdown rat plasma to their brain, adding to the body of knowledge that SePP1 primary function is selenium transport to the brain, with any antioxidant or cell viral response functions coming secondary [76].

Since the experiments in this paper are limited to HEK293T cells and not animal models, the complexities of SePP1 expression are limited, and it would stand to reason the need for TXNRD1 expression, or other antioxidant functioning selenoprotein expression including glultathione peroxidase would be favored over the expression of SePP1 in these cells, especially in context to cellular response against ZIKV infection. These could all be compounding reasons why SePP1 isoforms reported in the literature were not observed in even the HEK and HEK+Se conditions of these experiments in meaningful ways. The antigen for the antibody used for probe for SePP1 expression in all experimental conditions was the N-terminal region of the protein, which is present in all isoforms, so missing SePP1 isoform expression due to the antigen for the SePP1 antibody simply not being present in the shorter isoforms was not a variable [1-7]. These results also lend support to the proposed antisense interaction between ZIKV RNA and SePP1 mRNA in the host cell. The proposed antisense interaction between these two RNA species is predicted to occur upstream from the first UGA selenocysteine insertion site in SePP1 mRNA, meaning that if this antisense interaction between ZIKV RNA and SePP1 mRNA were to occur,

protein translation of any truncated isoforms of SePP1 would be highly unlikely [68-70, 77]. Additionally, viruses in the *flaviridae* family have the ability to replicate in the nuclease of the infected cell [64, 78, 79]. As stated, the antisense interaction proposed between ZIKV RNA and SePP1 mRNA could occur with the pre-mRNA form of SePP1 mRNA. This means that ZIKV viral RNA could be knocking down SePP1 pre-mRNA in the nucleus before it has the chance to enter the cytoplasm of the cell for protein translation. Additionally, a 2006 study showed that proteins essential for the ribosomal readthrough of the UGA stop codon to a selenocysteine codon, SECIS binding protein 2 (SBP2) and Specialized Elongation Factor (EFsec), both undergo nucleocytoplasmic shuttling through identification of export signals and immunohistostaining revealing both proteins colocalized within the nucleus [80].The unique ability of ZIKV RNA to enter the nucleus for replication and interact with pre-mRNAs like SePP1 mRNA could be a factor as to why no other isoforms of SePP1 are observed in ZIKV infected experimental conditions, in addition to ZIKV RNA having the unique ability to interfere with the selenocysteine insertion complex within the nuclear envelope as well as in the cytoplasm. This would lead to a near complete knockdown of SePP1 expression with no isoform expression, which is what is observed with these experiments. As stated, cellular expression of SePP1a is more abundant than other isoforms, so it is also possible that there was not sufficient SePP1 expression in general to identify other isoforms and a more sensitive test is needed to probe for any change of expression of additional SePP1 isoforms in this cell line, as it has been reported that immunoblot staining for the identification of SePP1 may not be the most reliable means of identification [75].

SePP1 expression in HEK lysate is highly visible for HEK conditions with no ZIKV infection, compared to HEK conditions with ZIKV infection, where SePP1 expression is hardly visual on

the western blot (**Figure 7, Panel B**). HEK average expression of SePP1 is reported at 43,500 for the positive control, and that increases to 80,100 when the cell environment is supplemented with 20nM sodium selenite. This increase in SePP1 expression from the control HEK lysate to HEK lysate with selenium supplementation is expected as several studies have reported an increase of selenoprotein expression with the addition of bioavailable selenium added to subject diets in environments were the subjects initially had a selenium deficient diet [77, 81-83]. Other studies with SePP1 knockout mice have shown that neurological symptoms like impaired motor coordination, audiogenic seizures, and brainstem neurode-generation can be avoided with appropriate levels of selenium added to the diet [6, 49, 84]. Going further, a 2004 study observed the progeny of SePP1 knockout mice, where the ability to synthesize specifically SePP1 had not only been diminished due to lack of dietary selenium, but completely negated. The SePP1 knockout mice were bred, and the resulting pups were also $SePP1^{-}$. The pups were further broken up into two experimental groups, one being allowed to nurse from a SePP1 knockout mother being fed dietary selenium and one nursing from SePP1 knockout mothers not being fed dietary selenium. The study found that when being allowed to nurse from the selenium supplemented mothers, the pups are able to circumvent the negative effects of the selenium deficiency caused by the inherited inability to express SePP1; these negative affects being neurodegeneration, seizures, and death [85]. Expanding these studies to the experiments described in this paper, sodium selenite, a bioavailable form of selenium, was added to non-ZIKV infected HEK cell culture medium for 3 full days and to ZIKV infected HEK cell culture medium 1 day after being ZIKV infected for 3 full days until the cells were harvested. With more selenium in the cellular environment, the cellular production of selenocysteine tRNA's is expected to increase, meaning that not only would there be a higher expression of SePP1 in the non-ZIKV infected selenium

supplemented HEK cells, but that a slight rescue of SePP1 expression would be observed in the ZIKV infected selenium supplemented HEK cells; as is the observation in several studies looking at the effects of selenium supplementation on selenoprotein levels.

SePP1 expression drops down to only 268 in ZIKV infected cells and increases to 416 once the ZIKV infected cells are supplemented with 20nM sodium selenite, however this slight 0.3% increase in SePP1 protein expression is not statistically significant or biologically relevant. In all experimental runs, SePP1 expression in HEK+ZIKV+Se experimental conditions as compared the SePP1 expression in HEK+ZIKV experimental conditions increased by an average of 197 (**Table A1**, **Appendix A**).

The lack of sodium selenite rescue of SePP1 expression in ZIKV infected cells is interesting especially when compared to the literature. An epidemiological study following the rate and severity hantaviral infection (RNA based family of viruses causing encephalitis that has shown to cause fatal acute neurological disease in mice) in mainland China has shown that people living in areas where the soil has deficient selenium levels have 6 times the likelihood of developing hantaviral related hemorrhagic fever with renal syndrome and a significant higher likelihood of mortality from the resulting condition [86]. The study saw that patients admitted to the hospital, once selenium supplemented, had increased levels of selenoprotein activity and plasma levels of selenium and alleviated symptoms [87]. Studies have indicated a high importance of TXNRD1 and glutathione peroxidase to the cell in regard to viral infection response, and have shown that even with selenium supplementation, that other selenoproteins in humans like glutathione peroxidase were prioritized and those concentrations were optimized in the cell before SePP1 levels were corrected upon viral infection, selenoprotein knockdown, or in cancer cells [58, 88]. Considering these observations from other labs in conjunction with the importance of selenium

recycling in the cell previously discussed, the conclusion could be drawn that other selenoproteins like TXNRD1 and glutathione peroxidase 1 were prioritized over SePP1 expression. Selenium is a limited resource and the mechanism for selenoprotein translation is already inefficient. TXNRD1 and glutathione peroxidase have stronger literary references to being tied to anti-viral response than SePP1, so it would stand to reason that even with the addition of sodium selenite, that the synthesis of these proteins would be prioritized by the cell in the presence of ZIKV infection.

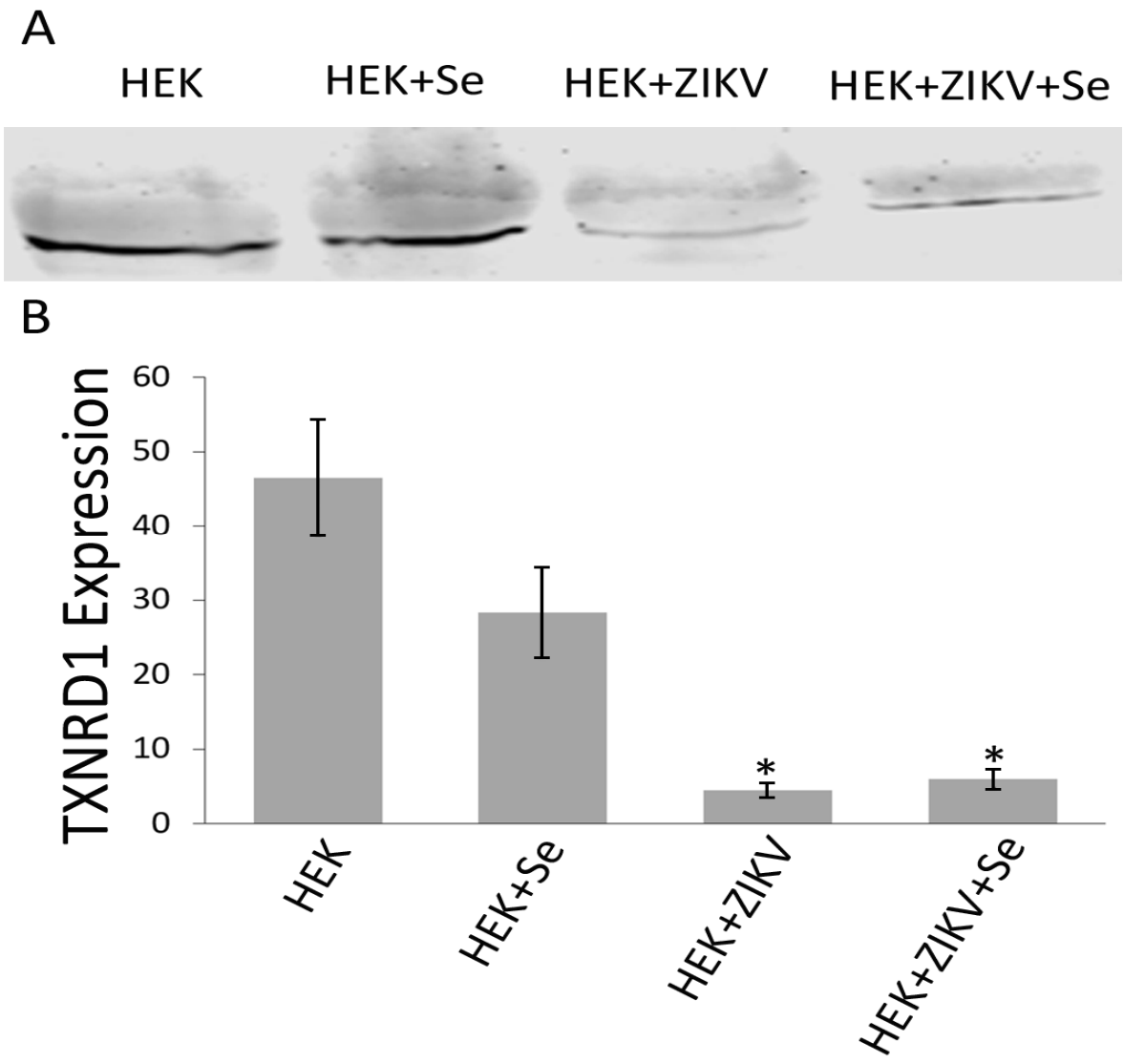


Figure 8: Western blot analysis of TXNRD1 expression in ZIKV infected HEK 293T cells. A. *Western blot of TXNRD1 protein expression under several experimental conditions* (with or without addition of 20 nM sodium selenite, either $ZIKV^{+}$ or $ZIKV^{-}$). Equal concentrations of protein were used per sample. Lanes are designated in the images. Results show: **A.** Anti-TXNRD1-labeled protein bands across experimental conditions at the 55 kDa level, which corresponds to the predicted mass of TXNRD1. About a 90% decrease in TXNRD1 protein expression is seen in ZIKV-infected cells. **B.** *Quantitative expression of TXNRD1 expression in experimental conditions* (Y axis shows fluorescence intensity units/1000). HEK TXNRD1 expression starts at 46,500, decreases to 28,300 for HEK+Se, drops to 4,480 for HEK+ZIKV, and increases slightly to 6,030 for HEK+ZIKV+Se. Error bars are ± standard error of the mean, * indicates a p-value ≤ 0.05.

An analogous western blot was performed with the same sample set probing for TXNRD1 expression across the experimental conditions: HEK cells with 20 nM sodium selenite supplemented into the cell media, HEK cells infected with ZIKV for 96 hours, and HEK cells infected with ZIKV for 96 hours and supplemented with 20 nM selenium for 72 hours. **Figure 8** (**Panel A**) shows expression of TXNRD1 across all experimental conditions at the 55 kdal line which corresponds to the molecular weight of TXNRD1 [89]. TXNRD1 expression in HEK lysate is highly visible for HEK conditions with no ZIKV infection, compared to HEK conditions with ZIKV infection; where TXNRD1 expression can hardly be visualized on the western blot. Image analysis data shows HEK positive control non-infected and non-supplemented TXNRD1 expression at 46,500 (**Figure 8**, **Panel B**). Surprisingly, this number decreases by 39.1% to 28,300 once sodium selenite is supplemented into the cell culture media (**Table A2**, **Appendix A**). Again, it is unlikely that this observation is a result of any cytotoxic effect of the sodium selenite, as the sample used to probe for TXNRD1 expression was the same sample used to probe for SePP1 expression. This indicates that something else has occurred. The majority of literature pertaining to TXNRD1 expression, as with SePP1 expression, shows that an increase of bioavailable selenium leads to an increase in TXNRD1 protein expression [81, 90]. One theory as to why this phenomenon was not observed with this experimentation is in the cell lines used for the experiments. HEK293T cells are from an immortalized human fetal kidney cell line; one study looking at zebra-seabream TXNRD1 expression when exposed to dietary selenium showed similar results to the ones displayed in **Figure 8** in the zebra-seabream kidney tissue samples [91]. The drop in TXNRD1 expression observed in zebra-seabream kidney cells when exposed to dietary selenium is similar in its quantity to the TXNRD1 drop witnessed in HEK293T cells when exposed to sodium selenite in this study. This lends credence to this drop

being a consistent and observable occurrence in TXNRD1 expression in kidney cell lines versus some other erroneous variable. Another study looking at the over expression of TXNRD1 in HEK293T cells found that the activity of TXNRD1 rather than the concentration of TXNRD1 is increased in these cell lines in the presence of environmental selenium supplementation, so while excess selenium has been supplemented into the HEK+Se experimental condition, it could only be affecting the efficiency of activity of TXNRD1 versus cellular concentration [92]. Providing additional insight to the slight decrease in TXNRD1 expression observed in the HEK+Se condition, several studies connect excess selenium with inhibitory effects on the TXNRD1 system that are correlated with growth, migratory, and differentiation ability of the cell [93-95]. These factors rather than any erroneous variable in the experiment design lend explanation to the TXNRD1 expression level in the HEK+Se experimental condition.

Similar to the observations presented concerning SePP1 expression in ZIKV infected cells, a dramatic decrease in TXNRD1 expression with ZIKV infection is seen. TXNRD1 expression in ZIKV infected cells drops to 4,480 which translates to a 90.4% decrease in TXNRD1 expression (**Table A2**, **Appendix A**). The observed decrease in TXNRD1 expression by 90.4% in ZIKV infected HEK cells strongly supports the hypothesis that ZIKV infection in the cell leads to a knockdown of TXNRD1 expression. The drop in TXNRD1 expression is significant when compared to the control, and while it was a large drop in protein expression, roughly 10% of TXNRD1 expression remained versus the near complete knockdown of SePP1 expression observed in ZIKV infected cells. In addition to the drop in TXNRD1 expression not being as severe, some positive effect on TXNRD1 expression in ZIKV infected cells with selenium supplementation was observed; which was not the case with SePP1 expression (**Tables A1** and **A2**, **Appendix A**). TXNRD1 expression improved to 6,080 in the HEK+ZIKV+Se condition,

which translates to a 3.40% increase in expression (with a total decrease in overall TXNRD1 expression being 86.8% from just HEK lysate) with sodium selenite supplemented into the cell culture media (**Table A2**, **Appendix A**). TXNRD1 selenium rescue in expression is minor but statistically significant, while any rescue observed with SePP1 expression is not. These observations regarding SePP1 and TXNRD1 expression in selenium supplemented ZIKV-infected cells are interesting, especially when comparing them to the observations regarding SePP1 and TXNRD1 protein expression in the HEK+Se condition. For SePP1, there seemed to be no observable rescue of protein expression with sodium selenite supplemented to the media, despite there being more protein expression in non-infected cells with sodium selenite supplementation. This is the inverse of what was observed with TXNRD1 expression in non-ZIKV infected vs ZIKV infected selenium supplemented condition. Despite there being an initial decline in TXNRD1 expression in the HEK+Se experimental condition, there is a notable increase from 4,480 in the HEK+ZIKV experimental condition to 6,080 in the HEK+ZIKV+Se experimental condition. While an 90.4% decrease in TXNRD1 expression is still substantial in the HEK+ZIKV experimental condition, it is not as large or as significant a drop in activity as compared to SePP1 expression in that same condition. This observation could be due to the cell prioritizing TXNRD1 and even glutathione peroxidase 1 expression over SePP1 expression as part of a cellular response to viral infection. Studies have also shown that when TXNRD1 expression in the cell is knocked down, that those cells experience strongly enhanced production of glutathione, which also increases cell susceptibility to selenium [95, 96]. A 2009 study demonstrated that in the presence of glutathione and selenium, 90% of ZIKV viral replication was knocked down with an ED_{90} dose of 2.5 mM of selenium in human cells [97]. Drawing explanations using these studies and the observation in these experiments, the less than predicted

increase of TXNRD1 expression in the HEK+ZIKV+Se condition could be a combined result of TXNRD1 knockdown, being caused by ZIKV infection, triggering a cellular response to increase glutathione production. This theoretical increase in glutathione production would be occurring in presence of excess selenium in the HEK+ZIKV+Se experimental condition; this is in addition to the cell already being more sensitive to environmental selenium due to the TXNRD1 knockdown caused by ZIKV infection. All of these factors could lead to the diminishing ZIKV viral replication while prioritizing glutathione expression in the cell versus TXNRD1. This is supported by the RT-qPCR data concerning ZIKV RNA presence in the HEK+ZIKV+Se (**Tables 2** and **3**). Future experiments looking at the cellular production of glutathione and the expression of glutathione peroxidase in relation to TXNRD1 expression in ZIKV infected cells is needed to further explore these connections.

Epidemiological data has shown that individuals with HIV-1 who are selenium deficient have a higher likelihood of the condition progressing to AIDS, indicating that there is a clear connection between selenium, selenoprotein expression specifically in regard to TXNRD1, and viral pathogenesis progression in the host [98]. Research has shown that TXNRD1 expression is up to 7 times higher in a developing fetus than in an adult, specifically in fetal multipotential neuronal progenitor cells, indicating that TXNRD1 expression is important to the developing and differentiating cells [63]. In addition, literature cites transmembrane protein AXL as an entry method into the cell by ZIKV, which is expressed in these cells [99, 100]. Studies have not only confirmed that these cells can be infected by ZIKV virus specifically, but additionally have shown that human fetal brain derived multipotential central nervous system progenitor cells, when infected with virus, aren't able to differentiate to astrocytes, dendritic cells or glial cells [54, 62]. These correlations between TXNRD1 expression and importance to the fetal developing

brain are additive to studies showing that when cells that do not express TXNRD1 or experience a significant decrease in TXNRD1 activity, they undergo an increased rate of apoptosis [13, 101-103]. Connecting these studies to the fact that cells not expressing TXNRD1 undergo apoptosis and the data this paper has shown demonstrating an 90.4% decrease in TXNRD1 expression in ZIKV infected cells, it can be extrapolated that ZIKV induced TXNRD1 knockdown in the fetus leads to apoptosis in fetal multipotential NPCs; this could be an additional compounding factor in the onset of microcephaly in ZIKV+neonates. Selenium deficiencies have already been shown to lead to increased development of dangerous secondary conditions the virus can cause in the inflected individual [104-107]. This suggests that addition of dietary selenium could aid in the circumvention of the onset of harmful secondary conditions. The experimentation presented here does not align fully with the literature however, with just a 3.40% rescue of TXNRD1 expression in ZIKV infected cells with the introduction of environmental selenium in the cell culture media. A limitation of the study was that selenium supplementation of the cell culture media was only tested at a 20nM concentration. This was chosen for this line of experiments based on other studies in the lab group involving the optimization of viral stop codon readthrough in the presence of dietary selenium [2]. If SePP1 expression is knocked down in the ZIKV infected cells, especially the placental cells, of a pregnant individual, then little to no selenium is being transported to the developing fetus. Not only is this a concern for developing neurons that need the micronutrient, but it is a concern for TXNRD1 expression in the fetal developing NPCs as well. Rescue of TXNRD1 expression in ZIKV infected individuals by dietary selenium could decrease apoptosis in multipotential fetal NPCs, thus reducing the risk of microcephaly onset. The western blot data in both **Figure 7** and **Figure 8** show ZIKV dramatically knocks down the expression of SePP1 and TXNRD1. These knockdowns in expression and what they mean for

the developing fetus, especially in terms of the developing neuronal cells, could be the perfect storm that leads to ZIKV congenital microcephaly. The non-effect selenium supplementation had on the protein expression of SePP1 and TXNRD1 despite extensive literature supporting the opposite conclusion is indicative of the need for future studies to look further into selenium supplementation, protein expression rescue, and a decrease in the onset of secondary conditions brought on by RNA based viruses like the cases observed with ZIKV and microcephaly.

2.4b RT-qPCR data

Table 2: Rt-qPCR results of amplified RNA species from RNA purified from experimental samples HEK, HEK+Se, HEK+ZIKV, and HEK+ZIKV+Se.

Experimental condition	Ct average	Standard deviation	P-value (comparing experimental conditions to HEK)	P-value (comparing HEK+Se and HEK+ZIKV+Se)	P-value (comparing HEK+ZIKV and HEK+ZIKV+Se)
SePP1 RNA levels					
HEK	20.58	±3.3			
HEK+ZIKV	17.99	±0.53	0.13		
HEK+Se	19.90	±0.90	0.72		
HEK+ZIKV+Se	19.14	±0.83	0.43	0.26	0.059
TXNRD1 RNA levels					
HEK	19.80	±2.2			
HEK+ZIKV	20.31	±4.0	0.83		
HEK+Se	19.23	±1.8	0.70		
HEK+ZIKV+Se	16.81	±2.121	0.097	0.13	0.17
ZIKV RNA levels					
HEK	28.12	±1.1			
HEK+ZIKV	25.59	±2.9	0.0055*		
HEK+Se	27.70	±0.43	0.79		
HEK+ZIKV+Se	26.83	±2.1	0.32	0.64	0.29

Primers designed for the amplification of SePP1 mRNA, TR1 mRNA and ZIKV RNA were used to amplify and determine original RNA levels of each amplicon for each experimental condition (**Table 1**). In columns 5 and 6, Ct values for the HEK+Se condition were compared against Ct values for the HEK+ZIKV+Se contition to determine if a statistically significant difference in the average Ct values occurred between non-ZIKV infected and ZIKV-infected HEK293T selenium supplemented cells. In columns 5 and 6, Ct values for the HEK+ZIKV condition were compared against Ct values for the HEK+ZIKV+Se condition to determine if a statistically significant difference in the average Ct values occurred between non-selenium supplemented ZIKV-infected and selenium supplemented ZIKV-infected HEK293T cells. P-values were generated using a Two sampled, 2-tailed t-Test assuming equal variance. Ct values for each condition for each experimental run are in **Appendix A**. * indicates p-value below 0.05.

Rt-qPCR data was collected on total RNA collected from all experimental samples for levels of ZIKV RNA, SePP1 mRNA and TXNRD1 mRNA. All experimental samples from which the data obtained were identical, with only different primers being used depending on what RNA amplicon was being targeted. In **Table 2**, the average Ct values of all amplicons of the different RNA targets for the rt-qPCR results from the purified RNA. All data in table was calculated from quadruplicate data points for each condition for each amplicon. The Ct value is inversely related to how much starting material (target RNA meaning either SePP1 mRNA, TXNRD1 mRNA or ZIKV RNA for these experiments) was in the original RNA sample before amplification. Meaning the lower the average Ct value, the more of that target RNA was in the original sample. In **Table 3**, relative quantification was analyzed by comparing a control Ct value (non-infected, non-selenium supplemented HEK lysate, $Ct_{control}$) to experimental Ct values (experimental conditions involving ZIKV infection, selenium supplementation, or both; $Ct_{experimental}$) with the formula $2^{\Delta Ct}$, where ΔCt was defined as $Ct_{control}$-$Ct_{experimental}$. When purifying the RNA from all samples, a cell count was performed on the cell pellets so that an equal number of cells from each sample would undergo RNA extraction. From this, equal amounts of total RNA starting material were added to the PCR wells to eliminate variables. ZIKV RNA was targeted as a means to show proof of viral infection and propagation in the experimental conditions and to test possible effects selenium supplementation had on ZIKV these variables in HEK 293T cells. Both mRNA for SePP1 and TXNRD1 expression were targeted to assess the effect ZIKV infection has SePP1 and TXNRD1 expression at the mRNA level and if selenium supplementation either negated or exasperated these effects.

Looking at the data for SePP1 mRNA amplification, the Ct average for HEK lysate only with no ZIKV infection or selenium supplementation is 20.58. As expected, RNA from the experimental

condition with selenium supplementation had a slightly lower average Ct value of 19.90 (**Table 2**) which translates to a 1.6 fold increase in SePP1 mRNA originally in the RNA samples from the HEK+Se experimental condition as compared to HEK lysate only. This falls in line with literature indicating that more selenium in the environment leads to an increase in selenoprotein expression as sodium selenite is a bioavailable form of selenium. As discussed, selenoproteins are formed with the translational insertion of selenocysteines through UGA stop codon readthrough. This amino acid can only be synthesized with materials coming from dietary sources of bioavailable selenium like sodium selenite, and an increase in the expression of these proteins is associated with an increase in dietary selenium, which by relations would indicate an increase in SePP1 mRNA in the cell as well. The predicted increase of selenoprotein expression was also seen at the protein level of SePP1 (**Figure 7**), with an increased expression of 84.1% (**Table A1**, **Appendix A**). The RT-qPCR data and the protein expression data match up and both support an increase in SePP1 expression at both the mRNA and protein level for the HEK+Se experimental condition.

Table 3: Relative analysis of initial RNA species present (SePP1 mRNA, TR1 mRNA and ZIKV RNA) in experimental samples versus control samples.

Experimental sample	$*2^{\Delta Ct}$ where $Ct_{control\ average} = Ct_{HEK\ average}$		$*2^{\Delta Ct}$ where $Ct_{control\ average} \neq Ct_{HEK\ average}$
SePP1 RNA levels			
HEK+ZIKV	6.0	**HEK+Se vs HEK+ZIKV+Se**	1.7
HEK+Se	1.6		
HEK+ZIKV+Se	2.8	**HEK+ZIKV vs HEK+ZIKV+Se**	-0.54
TXNRD1 RNA levels			
HEK+ZIKV	-0.70	**HEK+Se vs HEK+ZIKV+Se**	5.4
HEK+Se	1.5		
HEK+ZIKV+Se	8.0	**HEK+ZIKV vs HEK+ZIKV+Se**	11
ZIKV RNA levels			
HEK+ZIKV	5.8	**HEK+Se vs HEK+ZIKV+Se**	1.8
HEK+Se	1.4		
HEK+ZIKV+Se	2.5	**HEK+ZIKV vs HEK+ZIKV+Se**	-0.42

Fold differences in RNA species levels between non-infected/non-selenium supplemented HEK293T control samples and experimental ZIKV infected/selenium supplemented samples was calculated using the formula $2^{\Delta Ct}$, where ΔCt is defined as $Ct_{control}$-$Ct_{experimental}$. In column 2, $Ct_{control}$ is defined as the Ct average for RNA samples from the HEK only condition and $Ct_{experimental}$ is defined as the Ct average for any RNA samples from experimental conditions HEK+ZIKV, HEK+Se and HEK+ZIKV+Se. In column 5, $Ct_{control}$ is defined as either the Cr average for HEK+Se or HEK+ZIKV as indicated in column 4; and $Ct_{experimental}$is defined as the Ct average for HEK+ZIKV+Se. All Ct values used in formula calculations were derived from the calculated averages of all the Ct values for the relevant RNA species in each experimental condition. P-values derived in columns 3 and 6 were calculated using data from the triplicate experimental runs for each condition and for each primer set (**Appendix A**).

Surprisingly, the Ct average value for SePP1 mRNA from the HEK+ZIKV experimental condition was even less when compared to SePP1 mRNA levels from both the HEK lysate only condition and HEK+Se experimental condition, coming in at 17.99. This translates to 6.0 fold increase in the amount of SePP1 mRNA in the original sample from this condition as compared

to the HEK lysate only condition and a 4.4 fold increase when compared to the HEK+SE condition; indicating that with ZIKV infection, HEK cells produce even more SePP1 mRNA than both native HEK cells and HEK cells with sodium selenite supplementation. Despite the 99.4% knockdown of SePP1 protein expression that these experiments have shown to happen in ZIKV infected cells, more SePP1 mRNA is observed in the HEK+ZIKV experimental condition. This could support that the proposed RNA antisense interaction between SePP1 mRNA and ZIKV RNA is occurring, that the interaction either does not last long enough to trigger cell immune response to a double stranded RNA species in the scope of these experiments, or that since the nucleotide length of the interaction is predicted to be relatively short, that the cell could be mistaking it as RNA secondary structure, thus not targeting it for degradation anyway. This makes sense, because if the proposed antisense interaction occurred in such a way as to alert the cell of a double stranded RNA species, it would trigger an anti-viral cell immune response and thus would not be advantageous to the propagation of ZIKV RNA. It has been explained that the proposed antisense region between ZIKV RNA and SePP1 mRNA is highly conserved, so it would not make sense at an evolutionary level for the virus to have a conserved sequence in its genome that leads to host cell degradation of its RNA. In addition, evidence has been shown that not only can *flaviridae* viruses like ZIKV replicate in the nucleus of the infected cell, but that the UGA decoding complex for selenoprotein mRNAs, specifically in the case of SePP1 mRNA, also can occur within the cell nucleus. It is postulated that this is how SePP1 mRNA avoids non-sense mediated RNA decay in the cell [80]. This provides further evidence that the hypothesized interaction between SePP1 mRNA and ZIKV RNA is possible in terms of cellular localization of both RNA species, energetic favorability, and that temporary RNA secondary structures resulting from this interaction can avoid decay from cellular responses.

The increase in SePP1 mRNA with ZIKV infection could be due to a multitude of factors. A 2019 study correlates to an increase in SePP1 mRNA translation in cell with Hepatitis C infection and shows that SePP1 mRNA binds to RIG-1; an important protein involved in cellular immune response to viral infection. The study concluded that this interaction between RIG-1 and SePP1 mRNA caused a decrease in cellular immune response to Hepatitis C infection, thus increasing the proliferation rate of the virus [108]. This relationship between SePP1 mRNA induction, RIG-1 binding, and decline in cellular response to viral infection is interesting because ZIKV is part of the *flaviridae* viral family. This family of viruses incorporate a similar strategy by using subgenomic flaviviral RNA truncated from the 3'UTR region of their genomes to interact with RIG-1 to decrease cellular immune response [35, 38, 66, 109]. Interestingly, several reviews on the biosynthesis and regulation of selenoenzymes in mammals indicate that the translational mechanism of the ribosome reading through the UGA stop codon when translating a selenoprotein mRNA is the limiting factor in the expression of these class of proteins, and due to all that is involved in this translational insertion complex, a UGA readthrough as a selenocysteine codon is an inefficient phenomena [110]. Anything interfering with this already tendentious process would dramatically decrease the amount of selenoproteins produced by host cells regardless of how many times the interaction occurred, due to how involved the mechanism of selenoprotein expression is already. This could explain why an increase in SePP1 mRNA is seen in ZIKV infected cells. Even if the proposed interaction only happened briefly in the cell, it could be enough to pause the ribosome so that the SePP1 mRNA would get dislodged, resulting in it not being translated and leading to excess SePP1 mRNA in the cytosol whose translation was interrupted. However, if this one to be the case, an increase in the expression of truncated SePP1 proteins would be expected, even if they were non-functional proteins that had no

selenocysteines. As discussed with the observations regarding the SePP1 western blot, no such truncated proteins were observed. Viruses are known to increase reactive oxidative species (ROS) in the cells they infect, and selenoproteins are part of an antioxidant group of enzymes that helps mitigate these affects. This increase in ROS caused by viral infections like ZIKV lead to an imbalance between the normal levels of cellular ROS and the cellular antioxidant enzymes, triggering the cellular antioxidant defense system [50]. As SePP1 protein expression gets knocked down (**Figure 7**), the cell could prioritize SePP1 translation in an attempt to restore selenohomeostatsis, which could an additional reason why an increase of SePP1 mRNA is observed in the HEK+ZIKV experimental condition (**Tables 2** and **3**) [3, 77].

The average Ct value for SePP1 mRNA is most interesting in the HEK+ZIKV+Se condition. Here, the Ct value has increased from the HEK+ZIKV condition from 17.99 to 19.14; meaning that there is less SePP1 mRNA in the ZIKV infected samples with selenium supplementation samples than in the samples from the ZIKV infected only condition. There is still a 2.7 fold increase in SePP1 mRNA levels from the HEK only condition, but considering the large fold increase in SePP1 mRNA observed in the HEK+ZIKV condition and the 1.6 fold increase in SePP1 mRNA in the HEK+Se condition, a higher fold difference in the HEK+ZIKV+Se condition was expected from the HEK only condition. With what is known about selenium supplementation and selenoprotein expression, the prediction was that SePP1 mRNA levels in the HEK+ZIKV+Se condition would be higher than those in the HEK+ZIKV condition [77]. However, the opposite is observed, with the SePP1 mRNA levels being 0.45 fold less in the HEK+ZIKV+Se condition than the SePP1 mRNA levels in the HEK+ZIKV condition. This is a good indicator that the proposed antisense interaction between ZIKV RNA and SePP1 mRNA is happening in these infected cells versus some other mechanism of SePP1 protein knockdown

because while a significant rescue of SePP1 expression was not observed with the western blot data (**Table A1**, **Appendix A**), a decrease of SePP1 mRNA in the same experimental condition is observed. Since there is less SePP1 mRNA in the HEK+ZIKV+Se condition and slightly more SePP1 protein expression in the same condition, it could indicate that adding the extra selenium to ZIKV infected cells may not be encouraging more SePP1 protein expression at the mRNA level but rather making it easier for the SePP1 mRNA already transcribed to circumvent the proposed antisense interaction with the ZIKV RNA in order to be successfully translated.

The Ct average for the HEK lysate regarding TXNRD1 mRNA levels is 19.80, with the Ct average for the HEK+Se condition coming just slightly lower at 19.23, translating to a 1.5 fold increase in TXNRD1 mRNA levels from HEK lysate alone. This difference in TXNRD1 mRNA between the HEK condition and the HEK+Se condition is relatively the same one observed with the SePP1 mRNA levels between these conditions, with the difference between Ct values being 0.68 difference for SePP1 mRNA and 0.57 difference for TXNRD1 mRNA. This again coordinates well with literature that supports an increase in expression of selenoproteins in the presence of environmental selenium. What is interesting is that the average Ct value for the HEK+Se condition was lower than the positive control for TXNRD1 mRNA, indicates a higher level of TXNRD1 mRNA in the RNA sample from this condition, but the protein expression level of TXNRD1 for the HEK+Se experimental condition was 39.1% lower than that of the HEK alone condition (**Table A2**, **Appendix A**). So, the increase in TXNRD1 mRNA levels in the HEK+Se condition did not translate to an increase in TXNRD1 protein expression. Another lab studies the effects of selenium and TXNRD1 expression using a stability generated HEK293T cell line that overexpresses TXNRD1. They have reported that selenium supplementation of cell growth media leads to an overexpression of the TrxR1 gene rather than

translating into an increased protein amount [92]. This could aid in the explanation as to why there is an observed increase in TXNRD1 mRNA in the HEK+Se condition while a decrease in TXNRD1 protein expression. Regarding TXNRD1's role in cell homeostasis as well, it is widely reported that TXNRD1 overexpression in HEK 293T cells leads to the cells not being able to migrate or differentiate properly [90, 93, 96]. Based on this literary evidence, a possible conclusion regarding the increase in TXNRD1 mRNA in the HEK+Se condition and the concurrent decrease in TXNRD1 expression is that by the environmental selenium increasing the amount of TXNRD1 mRNA, it could have triggered some cell response calling for degradation of TXNRD1 protein.

Even more of interest, when looking at the TXNRD1 RT-qPCR data is when it is in comparison with the SePP1 RT-qPCR data. The average Ct value increases in the HEK+ZIKV experimental condition for TXNRD1 mRNA from 19.80 to 20.31, translating to a 0.70 fold decrease from HEK lysate. This means that HEK cells infected with ZIKV on average had less TXNRD1 mRNA than non-ZIKV infected HEK cells, which is a very large difference from the 6.0 fold increase of SePP1 mRNA in ZIKV infected cells that was previously observed. This is in conjunction with a smaller decrease in the knockdown of TXNRD1 in ZIKV infected cells. With SePP1, a 99.4% decrease is seen in protein expression along with the 6.0 fold SePP1 mRNA increase in ZIKV infected cells. With TXNRD1, only 90.4% of protein expression was knocked down and a 0.70 fold decrease is seen in TXNRD1 mRNA levels (although the differences in mRNA levels are not statistically significant) (**Table A4**, **Appendix A**). Applying the same logic used to explain the slight decrease in SePP1 mRNA and a slight increase in SePP1 protein was observed in the HEK+ZIKV+Se condition, the 0.70 fold decrease in TXNRD1 mRNA levels could be due to the fact that TXNRD1 mRNA is getting translated more readily in ZIKV infected

cells than SePP1 mRNA was. As previously stated, the literature demonstrates that cellular levels of TXNRD1 and glutathione peroxidase 1 are prioritized over SePP1 levels in the presence of virus. TXNRD1 and glutathione peroxidase 1 are also expressed at a higher rate than SePP1 is when the cell is undergoing stress [88]. Stated in the introduction, the ZIKV RNA:SePP1 mRNA antisense interaction spans an intron/exon junction, indicating that this interaction could be happening between ZIKV RNA and SePP1 pre-mRNA. Additionally, there is substantial evidence that points to the SePP1 mRNA:ZIKV RNA antisense interaction occurring within the cell's nucleus due to characterized traits of both *flaviridae* replication and SePP1 mRNA translation; this means that this antisense interaction could in theory occur with both mature and pre SePP1 mRNA species and within the nucleus. The predicted antisense interaction region between ZIKV RNA and TXNRD1 mRNA occurs on an intron, meaning that the chances of this interaction occurring are already less than with SePP1 mRNA species because ZIKV RNA only has the possibility of interaction with TXNRD1 pre-mRNA. In addition, not only is the predicted interaction shorter with ZIKV RNA and TXNRD1 mRNA than it is with SePP1 mRNA, but it is also less stable. Putting this together, not only does ZIKV have a shorter window of opportunity to interact with TXNRD1 mRNA, but that interaction is less stable and more likely to be targeted for degradation by native rnases. This could be why there is a less dramatic decrease of TXNRD1 protein expression in ZIKV infected HEK293T cells but also why there is less TXNRD1 mRNA in this condition as well.

However, that theory comes into question when the average Ct values for TXNRD1 mRNA in the HEK+ZIKV+Se experimental condition are considered. The average Ct for TXNRD1 mRNA in the RNA from the HEK+ZIKV+Se experimental condition was 16.81, which is 8.0 folds higher than RNA from HEK lysate alone. When comparing the Ct averages for TXNRD1 mRNA

levels in the HEK+ZIKV condition versus the HEK+ZIKV+Se condition, while the Ct average was markedly lower for the former (indicating more TXNRD1 mRNA in the RNA purified from the HEK+ZIKV+Se condition versus the HEK+ZIKV condition), the difference was not statistically significant when compared against the other, despite the Ct average difference actually being larger than the difference between the Ct average for the positive control and the HEK+ZIKV+Se condition. This increase in TXNRD1 mRNA in the HEK+ZIKV+Se condition could explain the 3.40% TXNRD1 increase in protein expression observed in the same experimental condition (**Figure 8**).

As expected, non-infected and non-selenium supplemented HEK lysate had an exceedingly high average Ct value for the ZIKV RNA amplicon target at 28.12 (**Table 2**). This indicates that it took over 28 of the 40 PCR cycles to get enough amplification of the ZIKV RNA target to get enough fluorescence to be visualized above the baseline. These results show that there was little to no starting material for amplification of this target to be successful in this sample and correlate to other studies looking at ZIKV Ct values and their negative controls [111]. A similar trend is seen is the non-infected HEK + Se lysate sample, where a 0.42 decrease in the Ct average for ZIKV amplicons was observed. This indicates a minimal increase in the amplification of theoretical ZIKV RNA as represented in **Table 2** where that slight decrease in Ct value translates to a 1.4 fold increase in theoretical ZIKV RNA being originally present in the sample. This data was not determined to be statistically significant, and any amplification observed in these two experimental conditions (HEK and HEK+Se) at this stage is likely due to non-specific binding of primers to the starting material versus any amplification of the ZIKV RNA target, as is reported with samples containing less than 105 RNA copies/mL; as the sensitivity of the RT-PCR assay is significantly decreased with Ct values approaching or ≥30 [112, 113]. In addition, the standard

deviation for the HEK+Se condition is considerably higher than the other standard deviations for the ZIKV amplicon data set; leading credence to the argument that any amplification is likely due to non-specific binding. ZIKV-infected HEK lysate shows a statistically significant decrease in Ct average value from the non-infected condition (**Table 2**) by 2.53 translating into a fold increase of 5.8 (**Table 3**) and keeping in line with average Ct values of previous ZIKV RNA/viral load studies utilizing RT-qPCR method. A Ct average of 25.60±1.138 is reported from the RT-qPCR data from serum containing 50,000 PFU/mL of ZIKV, and the observed Ct average value for the HEK+ZIKV condition is 25.59±2.9 indicating positive ZIKV infection with viral replication falling in line with the literature [111].

Observing the HEK+ZIKV+Se data sets, it's indicated that selenium added to the ZIKV infected HEK cells increases the Ct average of ZIKV amplicons by 1.24 translating to only a 2.5 fold increase in ZIKV RNA present in the sample as compared to the non-infected, non-selenium supplemented condition. This is 3.4 fold decrease in difference from the control when compared to the HEK+ZIKV condition, meaning that selenium supplementation overall decreased ZIKV viral RNA originally present in the sample. When directly comparing the average Ct values between the HEK+ZIKV and the HEK+ZIKV+Se experimental conditions, the difference was not found to be statistically significant (**Table 2**). When comparing the fold difference in these two conditions by treating HEK+ZIKV as the $Ct_{control}$ and the HEK+ZIKV+Se condition as the $Ct_{experimental}$, a 0.42 fold decrease in ZIKV RNA present from the HEK+ZIKV condition to the HEK+ZIKV+Se condition is observed. Multiple studies have listed inhibitory effects selenium supplementation has on various cellular stressors including viral infection and proliferation, so decrease in ZIKV RNA levels in ZIKV infected cells supplemented with selenium was the expected outcome. There is still a lack of protein rescue for both SePP1 and TXNRD1,

indicating that while selenium supplementation did reduce ZIKV RNA levels, it had no significant restorative effect for the selenobiochemisty of the cell. The experiments described in this paper did not use any free form amino acid supplementation in conjunction with the sodium selenite, other than the 5% glutamate added to the cell culture media, and no multivitamin or other supplements were added to the cell culture media. Studies citing a circumvention of negative viral affects and pathogenicity reported greater results when these variables were used concurrently with selenium supplementation [106, 114, 115]. Additionally, the dosage of sodium selenite used on the ZIKV infected HEK cells was significantly larger in the cited studies than in the one outlined in this paper, which could also explain why little protein rescue was seen despite seeing large differences selenium had on ZIKV RNA, SePP1 mRNA and TXNRD1 mRNA in ZIKV infected HEK293T cells. If selenium supplementation decreases the amount of viral RNA present, then the proposed antisense mediated knockdown of SePP1 and TXNRD1 protein expression would decrease, which could explain the slight increase in those levels of protein expression in the HEK+ZIKV+Se condition. In animal models, viruses have a documented reaction that can alter their genomes due to oxidative stress in selenium deficient or selenoprotein deficient cells. This creates a more pathogenic strain of the virus that is more virulent that the original infectant. It has also been reported that in glutathione peroxidase knockdown mice infected with Human Enterovirus, the viral genomic RNA change as compared to control mice undergoes mutations to become a more virulent strain [50, 116]. These observations reported by other researchers could be why a decrease in ZIKV RNA is observed in the HEK+ZIKV+Se condition, but no real protein rescue of TXNRD1 or SePP1 is witnessed.

2.5 Conclusions

We have displayed compelling evidence that ZIKV viral RNA has an affinity to interact with host selenoprotein mRNA coding for the expression of SePP1 and TXNRD1; and we have shown that this interaction is possible on the molecular level through a gel-shift assay using matching oligos to the predicted regions of interactions between the host cell selenoprotein mRNAs and the ZIKV viral RNA. The data put forward through these experiments clearly shows that knockdown of host cell selenoprotein expression of SePP1 and TXNRD1 occurs in ZIKV infected cells, with SePP1 expression dropping 99.4% and TXNRD1 expression dropping 90.4%. This effect of ZIKV infection seems to be unaffected by supplementation of selenium to the environment, with rescue of host selenoprotein expression not being at a level where it is biologically relevant. RT-qPCR data shows that SePP1 mRNA increases in the presence of ZIKV infection by about 6 fold, despite there being almost no noticeable protein expression. This adds further validation to the argument that ZIKV RNA is interfering with SePP1 mRNA. Addition of selenium to ZIKV infected cells affects TXNRD1 mRNA expression significantly by an increase of 8.0 fold. ZIKV RNA also decreases by almost half a fold. Given the cellular functions of SePP1 and TXNRD1, and the documented effects selenoprotein knockdown or selenium deficiency have on neuronal conditions/neuronal development, we have provided a case for the connection of ZIKV infection, selenoprotein knockdown, and the onset of microcephaly in ZIKV+neonates.

Acknowledgments

This research was supported by a gift from the Dr. Arthur and Bonnie Ennis Foundation, Decatur, IL, to E.W.T., and by High Point University research startup funds for J.A.R.

CHAPTER III

SELENIUM-DEPENDENT READTHROUGH OF THE CONSERVED 3'-TERMINAL UGA STOP CODON OF HIV-1 NEF

Lakmini S. Premadasa[1], Gabrielle P. Dailey[1], Jan A. Ruzicka[2] and Ethan W. Taylor[1]

[1] Department of Chemistry and Biochemistry, The University of North Carolina at Greensboro, Patricia A. Sullivan Science Building, PO Box 26170, Greensboro, NC 27402-6170, USA. [2] Department of Basic Pharmaceutical Sciences, Fred Wilson School of Pharmacy, High Point 10 University, High Point, NC 27268, USA.

3.1 Abstract

The HIV-1 nef gene terminates in a 3'-UGA stop codon, which is highly conserved in the main group of HIV-1 subtypes, along with a downstream potential coding region that could extend the nef protein by 33 amino acids, if readthrough of the stop codon occurs. Antisense tethering interactions (ATIs) between a viral mRNA and a host selenoprotein mRNA are a potential viral strategy for the capture of a host selenocysteine insertion sequence (SECIS) element (Taylor et al, 2016) [2]. This mRNA hijacking mechanism could enable the expression of virally encoded selenoprotein modules, via translation of in-frame UGA stop codons as selenocysteine (SeC). Here we show that readthrough of the 3'-terminal UGA codon of nef occurs during translation of HIV-1 nef expression constructs in transfected cells. This was accomplished via fluorescence microscopy

image analysis and flow cytometry of HEK 293 cells, transfected with engineered GFP reporter gene plasmid constructs, in which GFP can only be expressed by translational recoding of the UGA codon. SiRNA knockdown of thioredoxin reductase 1 (TXNRD1) mRNA resulted in a 67% decrease in GFP expression, presumably due to reduced availability of the components involved in selenocysteine incorporation for the stop codon readthrough, thus supporting the proposed ATI. Addition of 20 nM sodium selenite to the media significantly enhanced stop codon readthrough in the pNefATI1 plasmid construct, by >100%, supporting the hypothesis that selenium is involved in the UGA readthrough mechanism.

3.2 Introduction

Our group has previously proposed the possibility of viral RNA/host mRNA antisense interactions as a gain-of-function strategy for viruses, via the tethering of a host mRNA containing functional structural or sequence elements [2]. In that study, we focused on two viral genes, the nucleoprotein (NP) of the highly pathogenic Ebola Zaire strain (EBOV) and the HIV-1 nef gene, both of which terminate in highly conserved UGA stop codons, and both of which have, in close proximity to the UGA codon, extensive regions of antisense complementarity to the mRNA of an isoform of the mammalian selenoprotein thioredoxin reductase (TR3 in the case of the EBOV NP gene, and TXNRD1 in the case of HIV-1 nef). We proposed that these regions could provide the basis for "antisense tethering interactions" (ATI), and that in both cases the functional element being targeted by viral capture is the selenocysteine insertion sequence (SECIS) element located in the 3'-UTR of all human TR mRNAs [117]. This structural element is essential for translational recoding of in-frame UGA stop codons as selenocysteine (Sec), an amino acid whose presence is the defining feature of selenoproteins [118]. Thus it seems

unlikely to be a coincidence that the two genes in question, from very different RNA viruses, should both show evidence of antisense targeting of a human selenoprotein mRNA, that it is a TR isoform in both cases, in regions of the viral genome that prove to overlap or be within a few hundred bases on either side of a highly conserved in-frame UGA codon, recoding of which as Sec would lead to an extended selenoprotein isoform of the respective proteins (NP and nef), the sequence features of which are also conserved, even past the UGA stop codon.

In mammalian TR enzymes, the Sec residue encoded by the UGA is geminal to a Cys residue at the C-terminal end of the protein, the Cys-Sec pair forming the C-terminal redox center of TR enzymes [117]. As we have pointed out previously, this is strikingly similar to the situation in the HIV- nef gene, which has a conserved C terminal Cys residue, followed immediately by the UGA stop codon, thus potentially forming a TR-like Cys-Sec redox pair if the UGA codon could be translated as Sec [119].

Selenoproteins encoded in the genomes of bacteria, eukaryotes and archaea all encode Sec using the UGA codon, which otherwise, and far more commonly, serves as a stop codon [4]. To incorporate SeC into polypeptides, these organisms all use variants of a complex co-translational mechanism, a key feature of which is the involvement of a characteristic RNA stem-loop structure. In the mammalian system, this structure is the SECIS element, and is usually located in the 3' untranslated region (3' UTR) of the selenoprotein mRNA. Various protein factors, including SECIS binding protein2 (SBP2) and a special elongation factor, EF_{Sec}, work in conjunction with the SECIS element to bind $tRNA_{Sec}$ at the ribosome, suppress the termination factor, and thereby enable Sec incorporation at the UGA codon, preventing it from acting as a stop codon (Fig. 1, upper panel).

Given the universality of UGA codon use for Sec, it should not be surprising that viruses might also exploit this mechanism and benefit by the ability to synthesize alternate isoforms of certain viral proteins, e.g., by extending a protein via readthrough of a stop codon. The latter is actually a general strategy of some viruses, and recoding of UGA as Sec is only one special case of readthrough suppression, for which there are diverse mechanisms that can result in conventional amino acids being incorporated at any of the 3 stop codons, depending on the context. For example, murine leukemia virus (MLV) can only express its pol gene via translational readthrough involving the insertion of a glutamine at the position of a UAG stop codon that separates the gag and pol reading frames, an event that occurs at low but adequate efficiency (5%) [120]. This has evolved to give an appropriate 20:1 molar ratio in the quantity of gag structural proteins vs. the enzymes encoded in the pol gene.

In addition to a potential selenoprotein isoform of nef associated with its 3'-UGA codon, based largely on computational analyses of the genomic structure of HIV-1, we have also previously identified several regions overlapping known HIV genes that potentially encode selenoprotein modules expressed by ribosomal frameshifting, including one that is a homologue of glutathione peroxidase (GPx) [5], and one with similarities to the transcription factor NF-κB, the primary cellular on/off switch for HIV [119, 121]. The HIV-1 GPx was cloned and expressed using a human SECIS element, and was found to encode functional GPx activity [5], and to protect transfected cells against oxidant-induced apoptosis [41].

However, neither we nor others were ever able to identify a functional SECIS element encoded by an RNA virus, which engendered considerable skepticism in regard to our findings, e.g. regarding the HIV-1 encoded GPx. The missing piece of the puzzle was finally provided by our 2016 hypothesis [2], which is presented visually in the lower panel of Fig. 1. The central idea is

that a virus would not need its own SECIS element if it could hijack one from the host, via the mRNA tethering mechanism outlined above. Since SECIS elements have the ability to function even when carried by a separate mRNA [3], it is highly probable that the presence of a SECIS in a tethered mRNA could also serve to decode certain in-frame UGA codons as Sec on the tethering viral RNA, particularly if this interaction was the result of a coevolutionary process and structural and mechanistic features had been optimized by evolutionary selection.

Our preliminary support for this hypothesis in the case of the Ebola NP and HIV-1 nef genes was via a combination of computational analyses and confirmation of the antisense interactions at the DNA level, via gel shift assay of the specific fragments that were predicted to be involved in the core antisense interaction (what we called the ATI1 region), shown as interacting RNA secondary structures in Fig. 1 of Taylor et al [2]. The aim of the current study was to assess *in vitro* the predicted outcome of that interaction in the case of the nef gene, i.e., the potential for formation of a C-terminal extended isoform of nef, and to seek evidence of a role for TXNRD1 mRNA in the efficiency of that readthrough.

To that end, we designed three nef expression constructs based on a GFP expression vector that included the complete nef coding region inserted upstream of GFP. In between the nef and GFP coding regions were either one or both of the predicted antisense regions: ATI1, spanning the nef 3'-UGA codon (pNefATI1 construct), and ATI2, which is further downstream in the 3' end of the viral RNA (pNefATI2 construct, which contained both the ATI1 and ATI2 regions). These plasmids were designed so that in the "wild-type" pNefATI1 and pNefATI2 constructs, readthrough translation or bypassing of the nef 3'-UGA would enable translation of the downstream in-frame GFP domain. The third plasmid (pNefATIstop) was similar to pNefATI2, except for the presence of two in-frame stop codons (TAA and TAG) between the ATI1, ATI2

and GFP regions, to serve as a negative control by making GFP expression impossible even if UGA readthrough had occurred.

In brief, the results reported here exploit that set of nef ATI constructs, along with anti-TXNRD1 siRNA and appropriate control siRNAs, to demonstrate that readthrough of the nef 3'-UGA stop codon occurs with significant efficiency (~19%), is selenium-dependent, being more than doubled by the addition of 20 nM sodium selenite to cell culture media, and is significantly reduced, by about 67%, in the presence of an anti-TXNRD1 siRNA. These results unambiguously confirm the existence of an extended isoform of nef formed by translational stop codon readthrough and are consistent with the hypothesis that TXNRD1 may play a role in the readthrough via the proposed antisense mechanism.

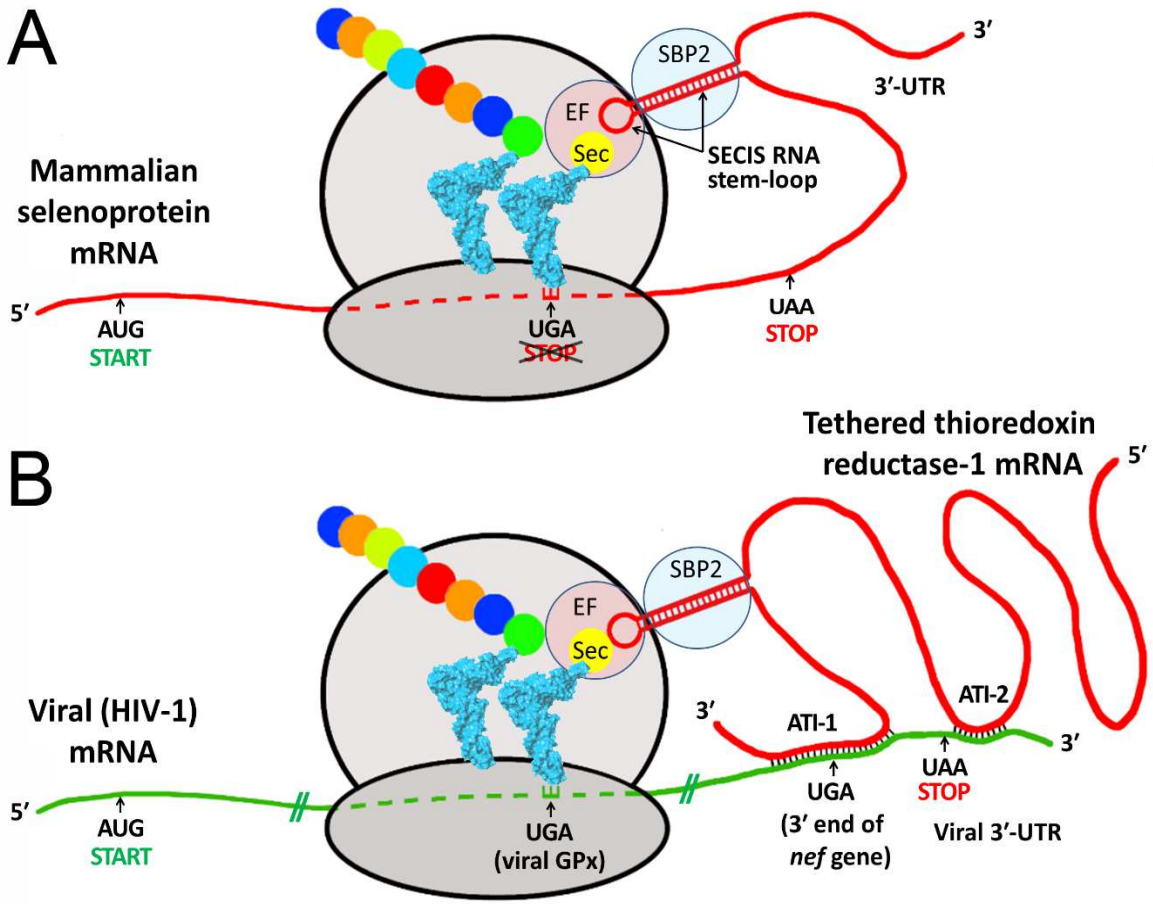


Figure 9: Proposed mechanism of Sec incorporation into viral proteins via hijacking of a SECIS element from a tethered host selenoprotein mRNA. (Reproduced from Fig. 3 of Taylor et al, 2016 [2]). Both panels show schematic ribosomes with two bound tRNAs, one carrying the Sec, the other a growing peptide chain shown as colored circles. The upper panel shows the established role of the SECIS element in the 3'-UTR region in the mechanism of insertion of Sec during mammalian selenoprotein biosynthesis [3, 4]. The panel below shows how the HIV-1 mRNA could hijack the host SECIS element via antisense tethering interactions (ATI) to decode UGA to synthesize viral selenoproteins such as the HIV-1 encoded GPx [5]. ATI-1 is a predicted interaction spanning the highly conserved 3' UGA codon of the nef gene. ATI-2 is a second shorter antisense region consisting of 13 consecutive bases near the end of viral mRNA.

3.3 Materials and methods

3.3a Engineering of HIV-1 nef expression vector constructs designed to monitor UGA stop codon readthrough

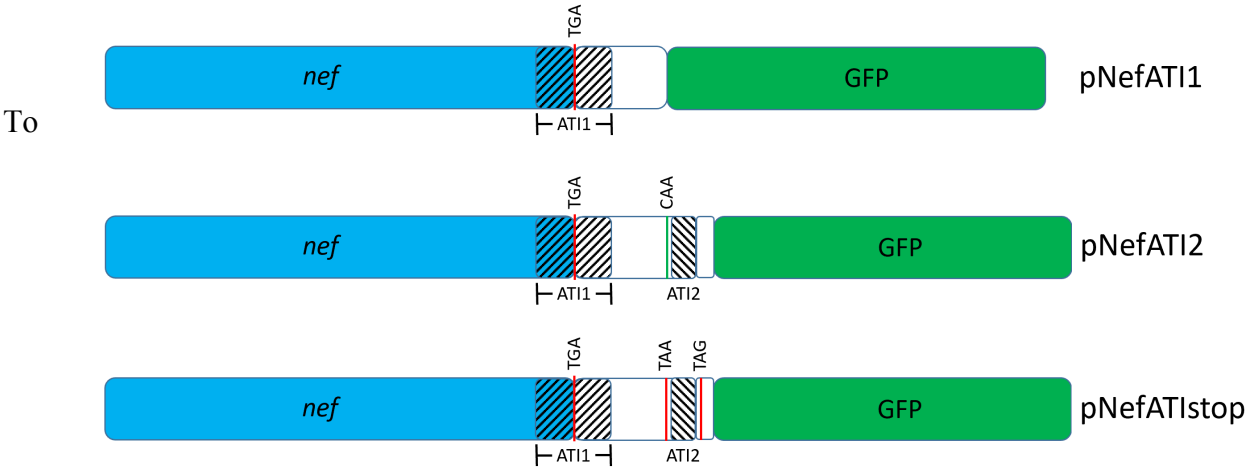


Figure 10: Schematics for the plasmid inserts used to assess nef 3'-UGA readthrough. Regios with antisense complementarity to TR1 mRNA are shown crosshatched as ATI1 and ATI2. Naturally occurring (TGA, TAA) or engineered (TAG) in-frame stop codons and a CAA mutant of the wild-type HIV TAA are indicated.

investigate the potential read-through of the UGA stop codon at the end of HIV-1 *nef* a set of three plasmid vectors was constructed, each containing a green fluorescent protein (GFP) reporter gene downstream of the complete HIV-1 *nef* coding region, including most of the 3'-LTR. The nef gene and LTR were obtained by PCR from the infectious molecular clone pNL4-3 (NIH AIDS Reagent Program 114; GenBank AF324493), and were ligated into the parent mammalian expression vector pAcGFP1-N3 (Takara Bio USA) such that GFP is expressed only if readthrough of the terminal UGA occurs. Primer design incorporated 5' NheI and 3' BamHI restriction sites into each of the three amplicons. To access GFP, an in-frame TAA stop codon at position 9504, between the two ATI regions, was mutated to CAA. The plasmids, shown schematically in Fig. 10, were 1) pNefATI1, which contained *nef* and only the first of the two putative ATI regions, 2) pNefATI2, in which the second tethering region was additionally

included, and 3) pNefATIstop, a double-stop negative control in which the stop codon at 9504 is retained and a second in-frame stop codon (TAG) is introduced just past the end of ATI2 but upstream of GFP, blocking formation of GFP even if readthrough of the UGA occurs. The primers (Integrated DNA Technologies, Inc, Coralville, IA; shown 5' → 3') included the same forward primer CAG GGC TAG CAA AGG ATT TTG CTA TAA CAT GGG TGG CAA G for all 3 constructs and reverse primers TGC AGG ATC CGA GGG CTC GCC ACT CC (pNefATI1), CCA GAG GGA TCC AGT ACA GGC AAA AAG CAG CTG CTT GTA TGC AGC ATC (pNefATI2), and GAG GGA TCC ACT ACA GGC AAA AAG CAG CTG CTT ATA TGC AGC ATC (pNefATIstop). Proper assembly of the vector constructs was verified by sequencing (Eurofins Genomics, Louisville, KY).

3.3b Selenium dependence of stop codon readthrough

To assess the influence of selenium on the readthrough of the terminal UGA codon in *nef* and optimize its concentration in the culture growth media, HEK 293T cells (NIH AIDS Reagent Program catalog number 103) were seeded (2×10^4/well) in a black clear-bottom 96-well plate (Costar 3603) and incubated for 24 hours in DMEM supplemented with 10% fetal bovine serum at 37 ^{0}C, 5% CO_2 atmosphere, after which the media was replaced with media enriched with sodium selenite (0, 20, 50, or 80 nM). After 48 hours the cells were transfected with pNefATI1 using lipofectamine 2000 (Invitrogen) by addition 50 μL of OptiMEM I (Gibco) containing 0.2 μg of plasmid DNA complexed with 0.3 μL of cationic lipid to each well. Incubation at 37 ^{0}C was continued for 12 hours and the media was replaced with fresh Se-enriched media. Four days post-transfection fluorescence was measured using a BioTek Synergy Mx microplate reader with excitation at 475 nm and emission at 505 nm.

3.3c Transfection of HEK 293T cells with pNefATI constructs

HEK 293T cells were transfected with the each of the pNefATI plasmids using the protocol described above, except that all incubations contained 20 nM Se (deemed the optimal concentration) after the initial 24 hour, Se-free period. The expression of GFP was followed using an EVOS cell imaging system. Fluorescence images were taken at 10 different well locations for each treatment and the average ± SD of the intensity was calculated using NIH ImageJ software[122].

3.3d Flow Cytometry Analysis

HEK 293T cells were seeded in a 6 well plate (5 x 10^5/well) in 1 ml of media enriched with 20 nM selenium and incubated for 12 hours. The cells were transfected with pNefATI1 or the parent pAcGFP1-N3 plasmid by addition of a complex of 0.5 µg plasmid DNA and 3.5 µL lipofectamine in 100 µL OptiMEM I. After 12 hours, the media was replaced with fresh media (20 nM Se) and the cells were incubated for 4 days at 37 °C. Cells were trypsinized, washed with 1% PBS, and resuspended in 400 µl of 0.1% BSA in PBS. GFP expression was evaluated using a BD FACSAria III flow cytometer with FACS DIVA version 6.1.3 software.

3.3e siRNA knockdown of TXNRD1 mRNA

TXNRD siRNA (Life Technologies 4390824) and the Ambion Silencer siRNA transfection II kit (Life Technologies AM1631) were used for the knockdown of TXNRD1 mRNA. Transfection of HEK 293 cells with pNefATI1 + siRNA was optimized by using varying volumes of the two transfection agents supplied in the Silencer kit, siPORT Amine and siPORT. All incubations used growth media enriched with 20 nM Se.

To initiate an experiment, siPORT Amine (0.6 µL + 25 µL OptiMEM I/well) was mixed with pNefATI-1 (0.2 µg + 25 µL OptiMEM I/well). Anti-TXNRD1 siRNA was reconstituted in nuclease-free water to a concentration of 20 µM and for each well 0.75 µL was diluted to a final volume of 25 µL in OptiMEM I. Aliquots (equivalent to 0.75 µL/well) of GADPH siRNA and scrambled siRNA (negative control) from the Silencer transfection kit were similarly diluted in OptiMEM I to a final total volume of 25 µL/well. The DNA/siPORT Amine complex solution (50 µL/well) and siRNA solutions (25 µL/well) were combined and 75 µL aliquots were placed in each well of a 24 well culture plate. Control wells contained 75 µL of OptiMEM with or without siPORT Amine and no nucleic acid. HEK 293 cells (5 x 10^5 in 425 µL media) were added to each well. The final concentration of siRNA was 30 nM. Plates were incubated at 37 °C for 72 hours. GFP expression was imaged using the EVOS system and fluorescence intensity was calculated from 10 images of each well using the NIH ImageJ software. An identical 24 well plate was prepared in tandem and used for qPCR analysis to quantify the siRNA knockdown of TXNRD1.

3.3f Quantitating TXNRD1 mRNA knockdown using RT-PCR

To confirm that the commercial anti-TXNRD1 siRNA was effective in our system, the cells in all treatment and control wells were harvested and processed to isolate total RNA (SV Total RNA Isolation System, Promega Corp., Madison, WI). cDNA was synthesized using 300 ng of RNA as template. As controls, a parallel set of reactions was run without reverse transcriptase and one reaction did not contain RNA template. A SuperScript first-strand synthesis system (Invitrogen) was used for cDNA synthesis. Power SYBR Green PCR master mix (Applied Biosystems) was used for qPCR which was performed on an ABI 7500 Fast PCR system (software version 2.3). The run conditions were 95 ^{0}C for 15 seconds, 58 ^{0}C for 15 seconds, and

60^0C for 60 seconds for each cycle. Expression of TXNRD1 mRNA in each sample was calculated relative to TXNRD1 expression in untreated cells. Values for TXNRD1 expression were normalized against expression values for the GAPDH housekeeping gene. The Pfaffl equation was used for calculation of relative expression.

3.3g Western blot probe for nef-GFP protein expression

1 mL of cold RIPA lysis buffer (10X Millipore, catalog number 20-188, diluted to 1X with deionized water) was added for every $5x10^6$ cells in the resulting cell pellets from each experimental condition. 10 uL of beta-mercaptoethanol, Halt™ Protease and Phosphatase Inhibitor Cocktail (100X, EDTA-free, Thermo Scientific™ catalog number 78441), and a 0.5 M EDTA solution (0.5 M EDTA in sterilized deionized water) was added to every 1 mL of cold RIPA lysis buffer prior to addition to the cell pellet. The cell pellet was resuspended in this solution and vortexed for 10 second intervals, three times each with 30 second intervals in between where the samples were incubated on ice. The samples were then sonicated for 15 second intervals, three times each with 30 second intervals in between each sonication where the samples were incubated on ice. The sonicator probe was cleaned using 99% ethanol in between each sample to ensure no contamination between the experimental conditions. The samples were then placed on a 95^0C heating block for 5 minutes before being spun down at 16.1xg at room temperature for 10 minutes. The supernatant was collected and stored at -80^0C until further use.

The Pierce BCA Protein Assay Kit (Thermo Scientific, catalog number 23225) was used to determine relative protein concentrations of each experimental sample. The assay was done to manufacturers specifications and visualized at 560 nm on a plate reader. Equal amounts of protein from each experimental condition were loaded into a 12% Mini-PROTEAN TGX SDS-PAGE Gel (samples contained 1X PBS, protein sample, and 3 uL 10X Loading Dye (Innovating

Science, catalog number IS5208)). 3 uL molecular weight ladder (SMOBio Enhanced 3-color*ExcelBand*™, catalog number PM2510) was added to one lane as a standard. The gel was set up in a BioRad Mini-PROTEAN vertical gel electrophoresis box with running buffer (10X Tris/glycine buffer, BioRad catalog number 1610732, diluted to 1X with deionized water) was added to the fill lines. An ice pack was placed in the gel box system and the setup was run at 80 V for 10 minutes and 115 V for 50 minutes. After running, the gel was removed from its cast and placed in deionized water along with 4 pieces of filter paper and 1 piece of 0.45 uM pore nitrocellulose membrane (BioRad catalog number 1620115) cut to the size of the gel. The wet sandwich was placed in the BioRad wet transfer casket and filled to the top of the set up with transfer buffer (1X running buffer, 20% methanol). An ice pack was placed inside the apparatus, and the transfer was allowed to run at 200mA for 1 hour and 15 minutes. Once the transfer was complete, the membrane was removed from the wet transfer casket and rinsed off once in deionized water before being blocked in 2% milk protein solution (2% w/v dried milk with 1X sterilized PBS) on a shaking table for 1 hour. The blocking solution was removed, and the blot was washed 3X for 5 minutes each with 1X sterilized PBS before addition of primary antibody (either monoclonal anti-GFP in rabbit (Sigma-Aldrich catalog number SAB2103123-100UL) or polyclonal anti-beta Actin (proteintech™ catalog number 20536-1-AP) depending on what protein expression was being measured). The blot was allowed to incubate with 10mL of 1:1000 dilution (in 1X sterilized PBS) of the primary antibody on a shake table at 4^0C overnight. Following this, the primary antibody was removed and the blot was washed an additional 3X with 1X sterilized PBS for 5 minutes a piece before 10mL of 1:5000 dilution (in 1X sterilized PBS) of secondary antibody (goat anti-rabbit IgG (H+L) Cross-Absorbed Secondary Antibody Alexa Fluor™ 700, Invitrogen catalog number A-12038 for GFP monoclonal antibodies and goat

anti-mouse IgG (H+L) Cross-Absorbed Secondary Antibody Alexa Fluor™ 700, Invitrogen catalog number A21036 for polyclonal beta-actin antibody) was incubated on the blot for 90 minutes on a shake table at room temperature. The secondary antibody was removed and the blot was washed as previously described before being placed in sterilized 1X PBS and visualized on gel imager Odyssey CLx (LI-COR) excited at 700 nm with emission measured at 720 nm.

3.4 Results

3.4a Readthrough of the HIV-1 nef 3'-UGA stop codon in HEK 293T cells transfected with nef-ATI-GFP constructs

GFP production from two versions of a wild-type HIV-1 nef expression construct designed to read past the conserved UGA stop codon into a downstream GFP domain (A and B in **Figure 11**) was compared to that from several controls: a negative control in which an in-frame TAA stop codon was inserted between the nef and GFP domains (C in **Figure 11**), the parent GFP expression plasmid as a positive control (D) and a second negative control consisting of untransfected cells (E). The relative GFP production calculated using ImageJ is shown in Panel F. Both the ATI1 and ATI1+2 constructs showed highly significant levels of stop codon readthrough relative to the negative control ATIstop construct, which was identical to the ATI2 construct except for two in-frame stop codons designed to prevent further translation into the GFP coding region (**Figure 11**). Differences between results for pNefATI1, pNefATI2 and the parent GFP construct (as 100% readthrough control) were all statistically significant at $P<0.001$. Unexpectedly, the inclusion of the second smaller anti-TXNRD1 region ATI2 in the construct lead to a decrease rather than the expected enhancement of readthrough (see Discussion). As expected, there was undetectable GFP expression from the untransfected control cells (E).

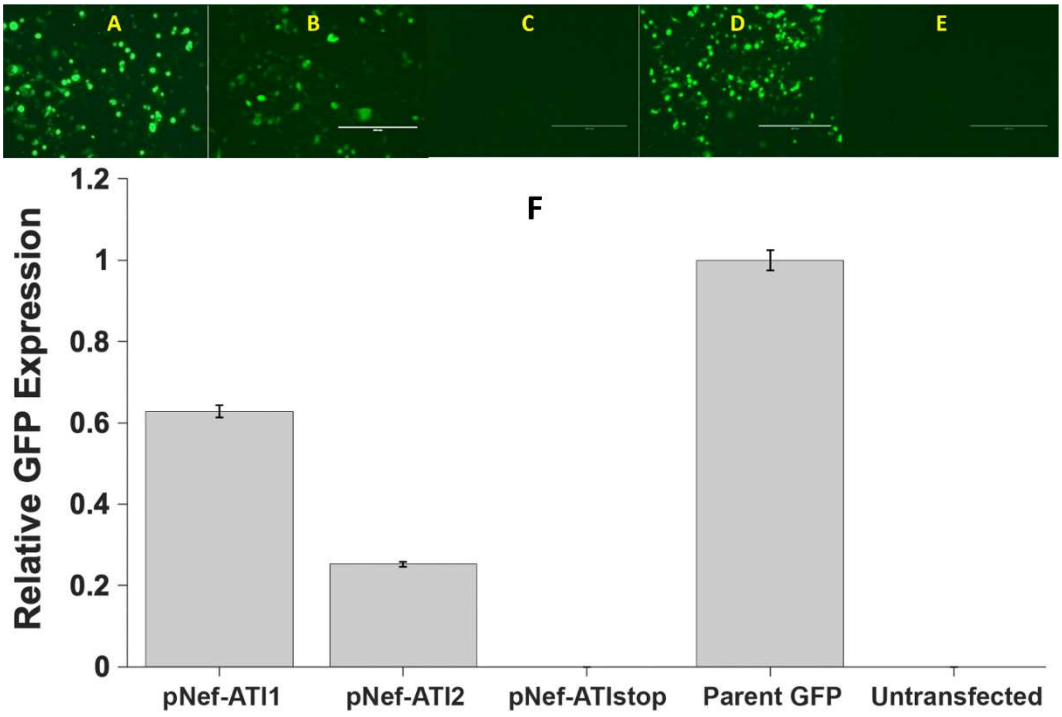


Figure 11: Readthrough of the HIV-1 nef 3'-UGA codon. The photomicrograph panels **A-E** show GFP fluorescence in HEK 293T cells, transfected with three different ATI plasmid vector constructs and controls. (A) pNefATI1; (B) pNefATI2; (C) pNefATIstop; (D) Parent GFP construct; (E) Untransfected cells; (F) Bar graph showing the GFP intensity of each transfection condition calculated using NIH ImageJ Software.

3.4b Selenium dependence of stop codon readthrough

Figure 12 shows the effect of varying the pretreatment selenium concentration on GFP expression from the ATI1 construct, which is a measure of the efficiency of UGA readthrough. HEK 293T cells were pretreated with different concentrations of exogenous selenium as sodium selenite, ranging from 0 to 80 nM final concentration, prior to transfection with the ATI1 plasmid. The results from Fig. 3 show that for pNefATI1, even without added selenite, GFP expression in standard culture medium is significantly higher than the untreated control ($p < 0.0001$). With additional Se supplementation, as seen in Fig. 4, the amount of readthrough product more than doubles going from 0 to 20 nM Se, for which the difference is also significant

(a 119% increase, $p < 0.0001$). However, GFP production levels for 20, 50 and 80 nM Se are not significantly different. Thus, at 20 nM of added exogenous Se, the mechanism appears to be saturated, as no further increase is observed at higher concentrations.

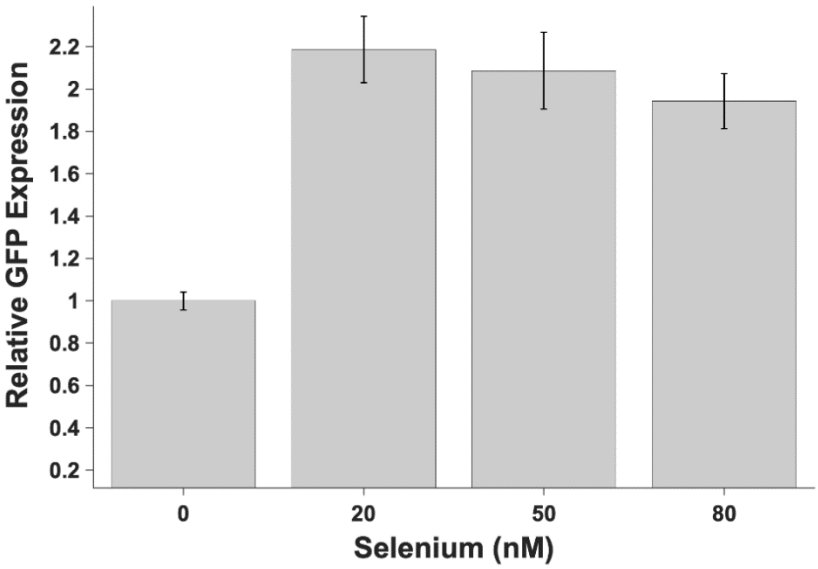
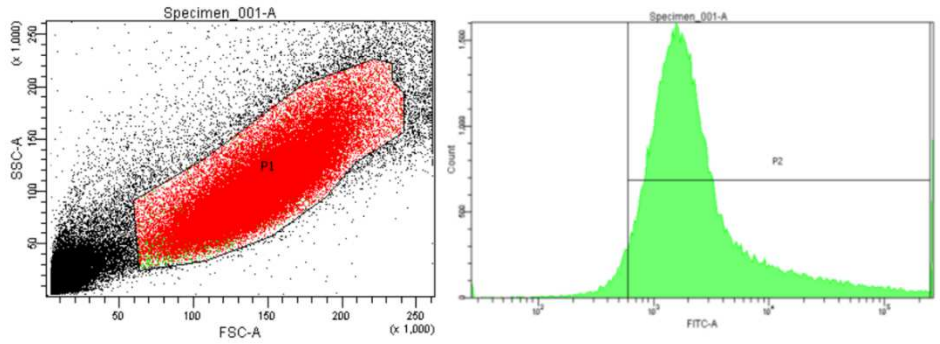


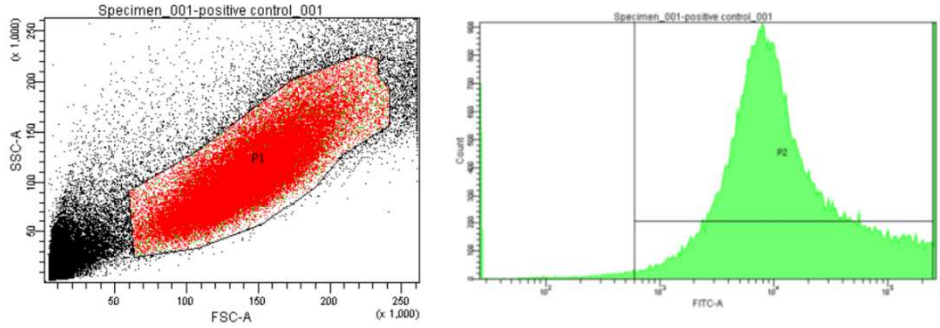
Figure 12: Added selenium enhances stop codon readthrough from the ATI-1 plasmid construct. Results show that supplementation of the basal media with additional selenium as sodium selenite has a significant effect on the stop codon readthrough, resulting in higher levels of GFP expression. However, the maximal readthrough was observed at the lowest level of selenium (20nM) and the readthrough did not change significantly at higher selenium concentrations. Even at the 20 nM concentration, addition of sodium selenite essentially doubles UGA stop codon readthrough relative to unmodified cell culture medium, as measured by GFP production.

3.4c Flow Cytometry Analysis of GFP expression from nef-ATI-GFP constructs vs. controls

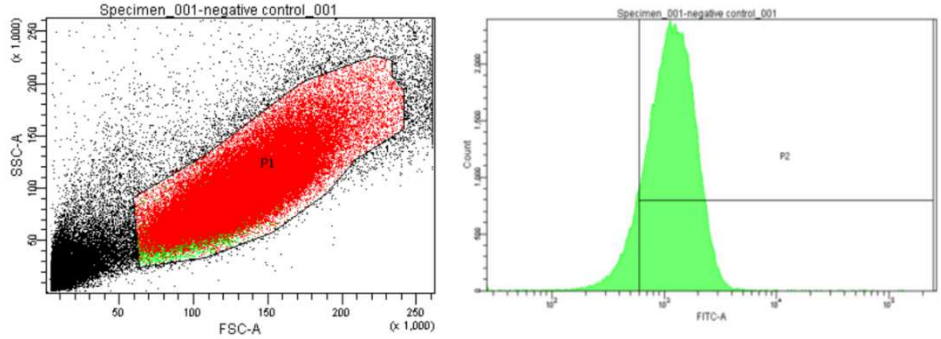
Flow cytometry was used to obtain a more accurate quantitative assessment of the nef UGA codon readthrough efficiency in the nef-ATI1 construct, relative to the parent GFP expression construct. As shown in **Figure 13**, the P1 population of cells transfected with pNefATI1 had a mean GFP fluorescence of 11,443 ± 1,242, as compared to the background of 1,304 ± 553 in untransfected control cells, and 55,076 ± 11,111 in the cells transfected with the parent GFP construct (a standard for 100% readthrough efficiency as it lacks any in-frame UGA codon).



Population	#Events	%Parent	FITC-A Mean	FITC-A rSD	FITC-A %rCV
P1	71,278	59.4	11,443	1,242	66.0
P2	67,428	94.6	7,658	1,195	62.3



Population	#Events	%Parent	FITC-A Mean	FITC-A rSD	FITC-A %rCV
P1	59,222	53.4	55,076	11,111	98.9
P2	50,515	85.3	26,074	8,066	82.4



Population	#Events	%Parent	FITC-A Mean	FITC-A rSD	FITC-A %rCV
P1	78,573	59.7	1,304	553	47.5
P2	70,520	89.8	1,368	513	41.4

Figure 13: Flow cytometry analysis of HIV-1 nef stop codon readthrough. Panel A (top) HEK 293T Cells transfected with pNefATI1 vector. FITC-A represents green fluorescence (GFP). The P1 population had a mean FITC-A of 11,443 ± 1,242. Panel B (middle) cells transfected with EGFP-N3 plasmid. The P1 population had a mean FITC-A of 55,076 ± 11,111. Panel C (bottom) untransfected cells (background). The mean FITC-A of P1 was 1,304 ± 553. Stop codon readthrough efficiency was thus 18.9% (see text).

3.4d siRNA knockdown of TXNRD1 mRNA decreases nef 3'-UGA codon readthrough

To assess the hypothesized role of thioredoxin reductase 1 mRNA in facilitating the UGA readthrough demonstrated in **Figures 11-13**, anti-TXNRD1 siRNA was used to knock down TXNRD1 mRNA levels, which we predict should lead to decreased GFP production from the nef ATI1 construct.

Figure 14 shows the results of a preliminary experiment used to determine the optimal transfection reagent and concentrations for the subsequent siRNA studies. This resulted in the choice of the siPort Amine at 0.6 µl per well as the transfection reagent for the siRNA TXNRD1 knockdown experiment shown in **Figure 15**.

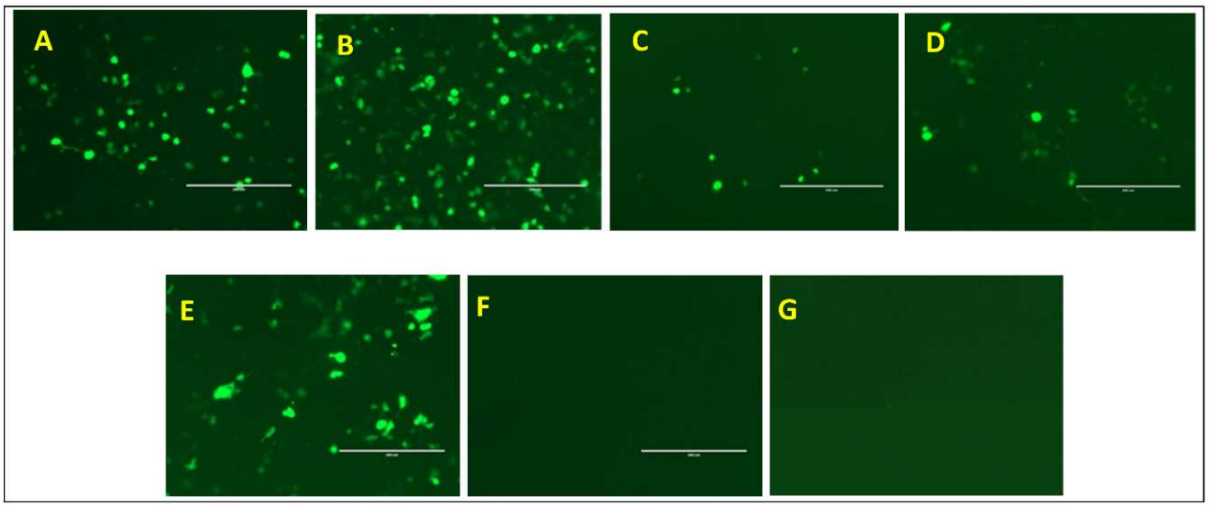


Figure 14: Selection of transfection reagent and optimization of the transfection reagent volume using HEK 293T cells transfected with pNefATI1 EGFP-N3 vector. EVOS GFP fluorescence images **A**, **B**, **C**, **D**, **E** and **F** are cells transfected with 0.3 ul amine, 0.6 ul amine, 0.15 ul amine, 0.5 ul NeoFX, 1.2 ul NeoFX and 0.15 ul NeoFx respectively. Image **G** represents untransfected cells.

In **Figure 15**, the effects of 3 different siRNAs on GFP production from the pNefATI1 plasmid are shown, as compared to pNefATI1 with no siRNA, and untransfected controls. The three siRNAs used were anti-TXNRD1, a random "scrambled" siRNA, and anti-GAPDH as a housekeeping gene control. The results show that the anti-TXNRD1 siRNA produces a

significantly greater decrease in pNefATI1-associated GFP production than either the scrambled or anti-GAPDH siRNAs. Compared to panel **A** (GFP production from pNefATI1 with no added siRNA), we observe a slight but non-significant decrease in GFP production in the presence of scrambled siRNA (panel **B**), a greater (~50%), significant decrease with the positive control anti-GAPDH siRNA (panel **C**; $p<0.001$), but the largest decrease of ~67% was with the anti-TXNRD1 siRNA (panel **D**; $p<0.0001$). There is no detectable GFP production from the untransfected control cells (Panel **E**). The average GFP values from the different conditions are compared in the bar graph in Panel **F**, from which the stated p values and percent decreases were calculated.

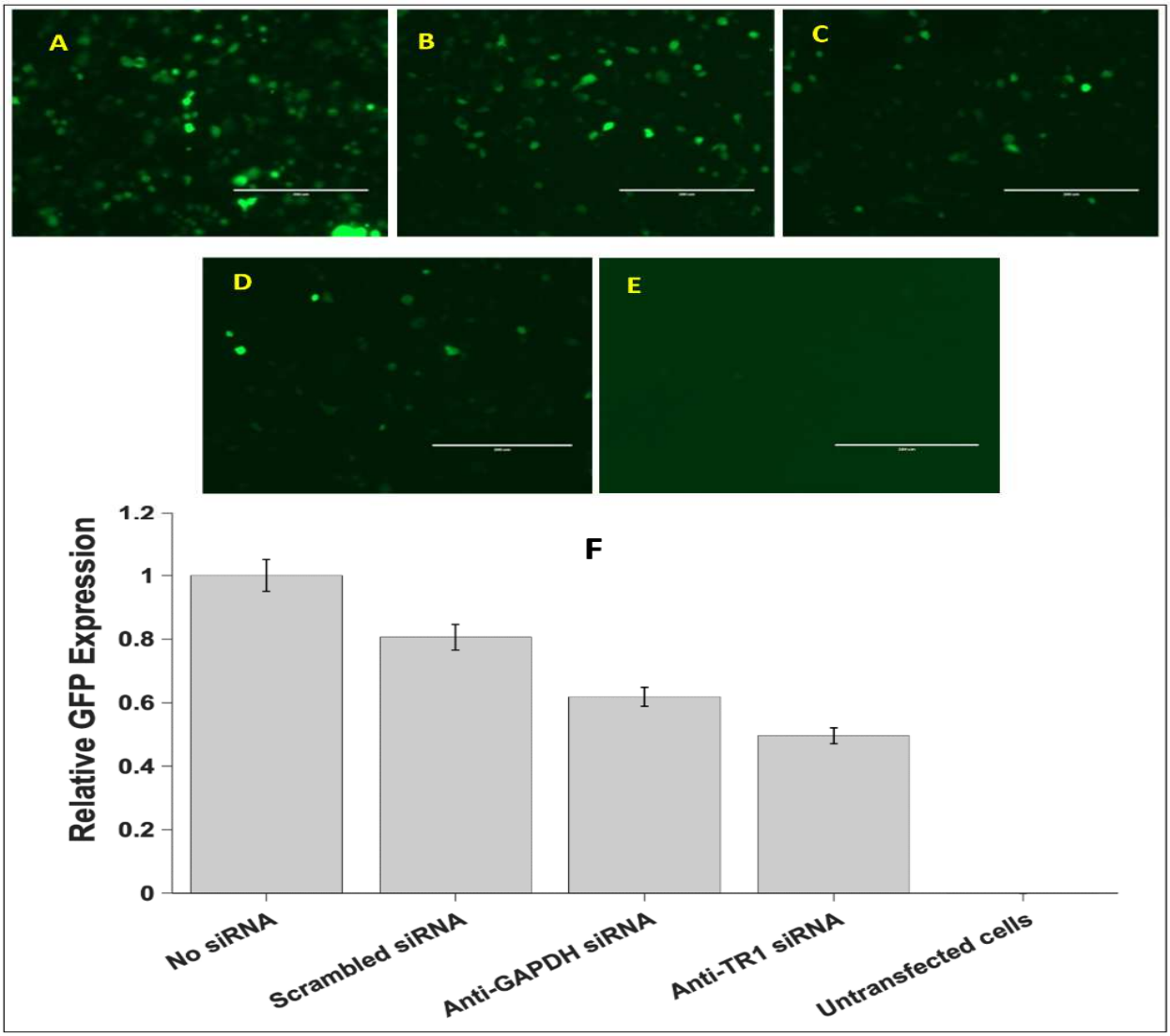


Figure 15: Effect of siRNAs on GFP production from the pNefATI1 construct. Panels A-E show EVOS microscopy images of each transfection condition: **A**. pNefATI1 with no siRNA. **B**. With added negative control scrambled siRNA. **C**. With positive control anti-GAPDH siRNA. **D**. With anti-TR1 siRNA. **E**. Untransfected cells. **F**. Bar graph showing average GFP expression from 10 images of each treatment calculated using NIH ImageJ software.

3.4e Validating TXNRD1 mRNA knockdown using RT-PCR

To validate the siRNA knockdown of TXNRD1 in the previous experiment (**Figure 14**), TXNRD1 expression in cells treated with the anti-TXNRD1 siRNA was assessed using qPCR, as compared to TXNRD1 expression in untreated cells, and cells treated with the scrambled siRNA, and cells transfected with the pNefATI1 construct alone. As shown in **Figure 16**, the presence of

anti-TXNRD1 siRNA lead to the largest and most significant decrease in TXNRD1 mRNA levels in treated cells, a decrease of about 28% under the conditions used ($p < 0.0001$). In cells treated with either the scrambled siRNA or pNefATI1 there was a small (~9%) and less significant decrease ($p < 0.01$) in TXNRD1 mRNA.

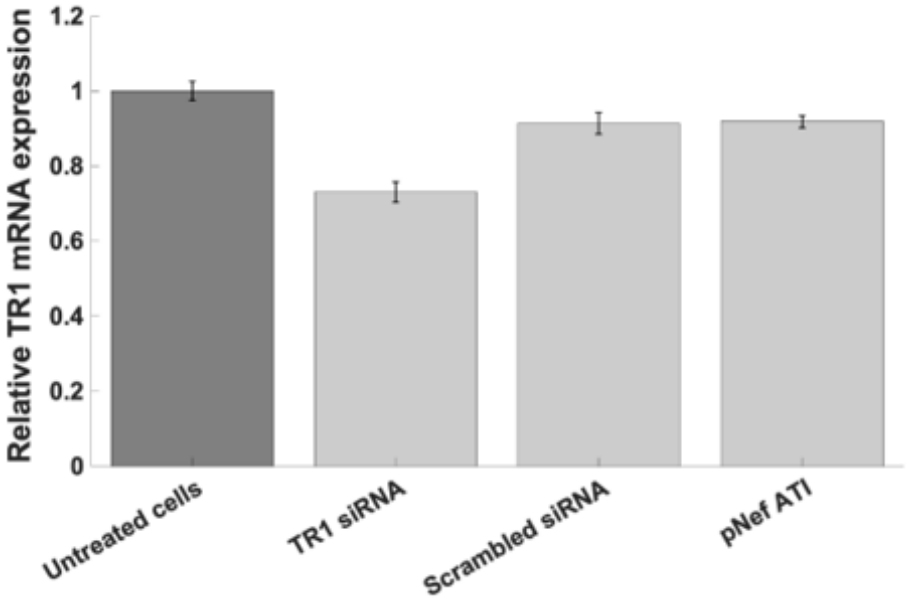


Figure 16: TR1 mRNA expression assessed by qPCR. Compared to untreated cells, an approximately 27% knockdown of TR1 mRNA relative to untreated cells ($P < 0.0001$) was observed in the sample treated with anti-TR1 siRNA. This confirms the knockdown of TR1 mRNA that may contribute to the decrease in nef 3'-UGA readthrough in the presence of anti-TR1 siRNA (**Figure 15**).

3.5 Discussion

The results of the current study may be summarized as follows:

- Using full length expression constructs of the HIV-1 nef gene containing downstream sequence elements extending as much as 130 nucleotides past the 3'-UGA stop codon, cloned upstream of GFP (as a reporter for translational readthrough), we have demonstrated that the UGA codon is suppressed at detectable efficiency, estimated by

FACS analysis to be about 19% by comparison to GFP production from the parent GFP construct.

- The nef expression constructs in question, pATI-1 and pATI-2, contain sequences of antisense complementarity to human TXNRD1, identified previously as regions of potential antisense tethering interactions (ATI) targeted by HIV-1. The dependence of nef 3'-UGA readthrough on TXNRD1 mRNA is supported by our finding that, in the presence of anti-TXNRD1 siRNA, there is a substantial and significant decrease (~67%, $p < 0.0001$) in UGA readthrough from the nef pATI-1 construct; this decrease is also significantly larger than the effect of either of the control siRNAs. This result supports our hypothesis that antisense tethering of TXNRD1 mRNA may contribute to the translational readthrough.
- When the Se concentration of the culture media was increased above basal levels by the addition of 20nM or greater sodium selenite, nef 3'-UGA readthrough increased by over 100%, supporting a role for selenium in the readthrough mechanism. However, our results do not address the question of whether this is due to the incorporation of Sec at the UGA codon, as hypothesized previously [2, 119], or by some other mechanism. A western blot probe of GFP expression was performed to address this question but gave conflicting results as we see over 100% more GFP expression in the pATI-1 condition when supplemented with selenium, but the protein product is the same molecular weight as GFP identified from the parent GFP plasmid protein expression.

These results strongly support our earlier predictions of translational readthrough of the HIV-1 nef 3'-stop codon, and a potential role for selenium [2]. However, they fall short of demonstrating that at least a subpopulation of translational readthrough products may contain

Sec at the location of the UGA codon. UGA is known to be a very "leaky" stop codon, and various amino acids with related codons (e.g. Trp, Cys, Ser) have been observed at this codon position in known examples, including mixtures of the various possible products, e.g., as reported by Chittum et al for beta globin [123].

This key finding is also consistent with experimental evidence for an extended isoform of nef obtained by researchers at the University of Rochester, NY, in which the nef extension was identified by immunohistochemical methods, in both post-mortem HIV+ brain slices and *in vitro* translated nef constructs from actual patient isolates (Dr. Benjamin Blumberg, personal communication).

The evidence presented here obtained using siRNA in regard to a possible role for TXNRD1 mRNA in the observed readthrough is positive, since there is a significant ~67% decrease ($p < 0.0001$) in GFP production from the pNefATI1 construct in the presence of anti-TXNRD1 siRNA, which is significantly greater than the reduction seen with either the anti-GAPDH control or the scrambled siRNA negative control; the decrease seen with the scrambled siRNA is not significantly different from no siRNA at all. The effect of the anti-GAPDH siRNA may simply reflect a decrease in overall cell activity produced by the targeting of an essential housekeeping gene.

We verified TXNRD1 knockdown by this commercial anti-TXNRD1 siRNA using RT-PCR, which in our hands showed only a 27% decrease in TXNRD1 mRNA levels. The manufacturer claims that the knockdown achieved with this siRNA should be greater, about 70%, which would be more consistent with the potency we observed for its effect on inhibition of GFP production from the pNefATI1 plasmid (a 67% decrease); however, it is possible that only a 27% decrease in TXNRD1 mRNA could have a somewhat amplified effect on a downstream system.

One important observation is that pNefATI1 produced only a minimal (9%) knockdown of TXNRD1 mRNA, similar to that produced by the scrambled siRNA. This is despite the fact that pNefATI1 *does* express an mRNA with a region of antisense complementarity to TXNRD1 mRNA. However, antisense and miRNA type interactions do not necessarily lead to degradation of a target mRNA, which depends on a high degree of complementarity of the 2 RNAs; when the antisense interaction is not perfect, it can be the basis of ribosomal interference with the target protein synthesis, without mRNA degradation [124]. Thus our results suggest that the antisense targeting of TXNRD1 mRNA by the HIV-1 nef 3'-region, in addition to serving a possible tethering function for TXNRD1 SECIS capture [2], may also lead to inhibition of TXNRD1 protein synthesis in infected cells, but not via degradation-mediated downregulation of TXNRD1 mRNA.

One unexpected result was that the readthrough efficiency was LESS when both anti-TXNRD1 ATI regions were present in the nef construct, as opposed to just ATI1 alone. Because in nef ATI1 the 3'-UGA is right near the center of an ~40 residue sequence that is complementary to a region of TXNRD1 mRNA, during translation up to and past the stop codon, which we have now shown does occur (Figs. 3, 4 and 5), ATI1 would have to become unwound as that region of the nef mRNA enters the ribosome. Thus, we suspected that the second downstream ATI region, though smaller (only 13 base pairs in length), might be important to keep the TXNRD1 mRNA and its SECIS element in proximity to the ribosome. However, it is also possible that the dynamics of the engagement of the TXNRD1 SECIS to the ribosome are such that they could continue to function in the recoding of the heterologous viral UGA, for at least a short time after the unwinding of ATI1. So ATI2 may not be particularly important. But it is also possible that the reason for the decrease in readthrough efficiency of the construct that includes ATI2 is

simply due to the fact that, in order to engineer this construct to include the ATI2 region, while at the same time permitting readthrough into the downstream GFP region, a region of viral 3'-RNA that normally has non-coding roles must be translated. Thus, codon usage in this region may be non-optimal, affecting translation efficiency in our ATI2 construct. In contrast, readthrough of the 3'-UGA, which we have now observed in transfected cells, would by necessity lead to C-terminal extension of the nef coding region by up to 33 amino acids, so that extended nef codon usage in the entire region spanning ATI1 has presumably been optimized for efficient translation during viral evolution. That is not the case for the region spanning ATI2, due to the TAA stop codon at its 5' end. In addition, the noncoding far 3'-end of the HIV-1 mRNA contains a number of RNA hairpin structures with diverse regulatory functions, the unwinding of which might slow translation around the ATI2 region.

An alternative explanation for GFP expression from these constructs could involve initiation of GFP protein translation at an internal start codon. Although there is a candidate methionine in the N-terminal region of the GFP coding sequence, there are no in-frame Met codons past the 3'-UGA of nef in any of the included viral sequences. Hence, the possibility of internal initiation of GFP synthesis fails to explain our observations, because if it was happening from the GFP internal Met, GFP production would also be observed with the pNefATIstop control construct, which is not the case.

Finally, it must be emphasized that the current results showing a link between selenium biology and the expression of an isoform of a gene (nef) that is central to HIV-1 pathogenesis should be interpreted in the light of an extensive body of evidence linking HIV disease progression and outcome to selenium status. Observations of selenium abnormalities in AIDS patients date back almost as far as the discovery of the virus in the 1980s (e.g., [125, 126]). Later studies have

linked HIV-associated mortality risk to serum selenium levels [105, 127, 128], and demonstrated various clinical benefits for selenium supplementation, either alone or as part of a micronutrient formulation [114, 115, 129].

Nor is HIV-1 an exceptional case in this regard. A number of predominantly RNA viruses have been shown in animals or humans to have either increased frequency and/or severity of cases in low selenium areas or individuals, or some measurable clinical or survival benefit from selenium supplementation (as reviewed, [14, 130, 131]). The most recent example is seen with the COVID-19 pandemic, for which Zhang et al. have now demonstrated a highly significant association between the outcome of SARS-CoV-2 infection and previously documented regional selenium status in Chinese cities [132].

Based on our results to date, and as proposed previously [21], we predict that antisense interactions with TR isoforms will be identified for a number of other RNA viruses, and that this may prove to be a general mechanism for RNA viruses to enhance RNA levels for virus production, by partially inhibiting the conversion of ribonucleotides to 2- deoxyribonucleotides. Because ribonucleotide reductase uses the thioredoxin system as a hydrogen donor, antisense knockdown of TR isoforms could tip the balance in favor of RNA production, by inhibiting the regeneration of reduced thioredoxin. This biochemical role of selenium in mammals may prove to be a significant factor contributing to the increased virulence of some RNA viruses in selenium-deficient hosts.

3.6 Conclusions

We have provided compelling evidence that sequences within the HIV-1 nef mRNA, with its proximal downstream elements, are sufficient to permit readthrough suppression and thereby

extension of the nef protein into the downstream region. The only non-native element introduced into the pNefATI1 construct was mutation of the TAA stop codon at the end of that extension region, in order to permit translation of the downstream GFP reporter gene. We have also demonstrated that addition of a low concentration (20 nM) of selenium as sodium selenite can increase the readthrough efficiency by over 100%, and that knockdown of TXNRD1 mRNA can reduce the efficiency of the stop codon readthrough by about two-thirds. These findings are consistent with a role for TXNRD1 mRNA and selenium-based cellular mechanisms in the readthrough event, but further studies will be required before we can say that virus/host mRNA tethering reactions are definitely involved in these observations – however, it has been strengthened as a hypothesis.

CHAPTER IV

DIRECT DETECTION OF CONSERVED VIRAL SEQUENCES AND OTHER NUCLEIC ACID MOTIFS WITH SOLID-STATE NANOPORES

Komal Sethi[1], Gabrielle P. Dailey[2], Zahid, Osama[1], Ethan W. Taylor[2], Jan A. Ruzicka[3], and Adam R. Hall[1,4#]

[1]Virginia Tech-Wake Forest University School of Biomedical Engineering and Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27101, USA. [2]Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, NC 27402, USA. [3]Department of Basic Pharmaceutical Sciences, Fred Wilson School of Pharmacy, High Point University, High Point, NC 27240, USA. [4]Comprehensive Cancer Center, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.

4.1 Abstract

Rapid and reliable recognition of nucleic acid sequences is essential to a broad range of fields including genotyping, gene expression analysis, and pathogen screening. For viral detection in particular, the capability is critical for optimal therapeutic response and preventing disease transmission. Here, we report on a novel approach to the detection of identifying sequence motifs within single-strand DNA and RNA viral genomes based on solid-state nanopores. By designing DNA oligonucleotide probes with complementarity to target sequences within a target genome, we establish a protocol to yield affinity-

tagged duplex molecules the same length as the probe only if the target is present. The product can subsequently be bound to a protein chaperone and analyzed quantitatively with a selective solid-state nanopore assay. We first use a model DNA genome (M13mp18) to validate the approach, showing successful isolation and detection of multiple target sequences simultaneously. We then demonstrate the protocol for the detection of RNA viruses by identifying and targeting a highly conserved sequence within human immunodeficiency virus (HIV1-B).

4.2 Introduction

Viruses are nanoscopic particles containing genetic material that must invade living organisms and take over their cellular machinery to replicate. In support of this mechanism, they can be both highly transmissible and rapidly mutating, enabling them to evolve quickly. From the identification of Yellow Fever virus as the first reported human virus in 1901, viruses have been a consistent threat to human health. Indeed, three to four new species of human virus are identified every year accounting for over two-thirds of new human pathogens [133-136]. In recent history, diverse diseases have emerged from viral vectors including human immunodeficiency virus (HIV), Ebola, avian influenza (H7N9, or bird flu), Zika, and middle east respiratory syndrome (MERS) [137, 138]. Enhanced global movement has further exacerbated the problem of rampant transmission, increasing the likelihood of global pandemics like the one caused by SARS-CoV-2 beginning in 2019 [139].

Viral detection capabilities are critical for combating the spread of such diseases, providing both accurate diagnosis for early treatment planning and a means of performing contact tracing for containment [138]. While a broad range of diagnostic approaches exist for this purpose, current technologies still encounter significant challenges. For example, tests incorporating polymerase

chain reaction (PCR) – including quantitative PCR (qPCR) and reverse transcription PCR (RT-PCR) – are considered gold standards in viral diagnostics [138, 140]. While these techniques are highly sensitive, however, they can also feature a risk of contamination and bias, commonly rely on expensive optical systems for analysis, and are often slow. Indeed, the typical hours-long cycling necessary for most PCR assays is a critical factor limiting throughput and increasing time-to-answer [141, 142]. Alternatively, immunoassays are also used extensively to identify protein markers associated with viral exposure. While high sensitivity can be achievable through several measurement modalities including fluorescence, radiolabeling, and colorimetric (biochemical), immunodetection also relies on expensive reagents and can suffer from cross reactivity, resulting in false negatives or false positives [143-146]. Finally, recent developments have been made in the application of next-generation sequencing (NGS) and mass spectrometry to viral diagnosis. While promising, these approaches remain nascent, hinging currently on expensive instrumentation and computationally-demanding analyses [138, 146].

An emerging platform for biomolecular analysis is the solid-state (SS-) nanopore [147, 148]. Briefly, a SS-nanopore is comprised of a thin, insulating membrane supported by a silicon chip and featuring a single, nanometer-scale pore. When the membrane is used as a partition between reservoirs of electrolyte solution, the application of a voltage bias across it generates an electric field that can facilitate the electrokinetic movement of charged molecules through the opening. Each translocation temporarily interrupts the measured ionic current and results in a signal (or 'event') that can be exploited for biosensing and molecular evaluation. SS-nanopores have been employed extensively to identify nucleic acids, proteins, biological nanoparticles (e.g. exosomes), and synthetic nanoparticles [149-157]. Recently, the platform has also been applied to viral diagnostics with several reports describing detection or analysis of whole virions,

including Paramecium Bursaria Chlorella virus (PBCV-1), Hepatitis B, HIV, and influenza [158-165]. However, probing whole viruses does not explicitly provide a route to viral discrimination without *a priori* knowledge of the pathogens present in a fluid and therefore the translational potential of these approaches has been narrow. The alternative approach of detecting specific viral biomarkers has been explored with the platform as well. For example, Niedzwiecki, *et al.* studied nucleocapsid protein 7, a protein constituent specific to HIV, by probing the effect of binding to a synthetic RNA on its translocation dynamics. However, the use of SS-nanopores to detect nucleic acid (sequence-based) viral bioindicators has been more challenging. This is at least in part because the platform itself lacks intrinsic selectivity: since all passing molecules produce a signal conventionally, differentiation of particular target sequences is difficult [166, 167].

To address this gap, our group has established a novel approach to detect and quantify targeted nucleic acid sequences [168]. Briefly, the two-component assay employs a short nucleic acid fragment (<~250 bp) featuring a single biotin tag and a small, globular protein (54 kDa monovalent streptavidin, or MS) [169, 170]. When each is introduced individually to a SS-nanopore of appropriate size (typically ~7.5-11 nm diameter), they do not yield significant events due to their small size and rapid transit time (Figure 1a-b). On the contrary, when bound together to form a nucleoprotein complex, the larger structure interacts sterically with the pore walls to slow the translocation speed and produce translocation events (**Figure 18C**). We previously extended this basic assay to allow the detection of nucleic acid sequences by exploiting the observation that events are generated by protein-bound duplex constructs but not single-strands bound to protein [171]. Through introduction of a biotin-labeled DNA probe to a mixture of single-strand sequences, duplex formation is possible only when antisense molecules

are present, yielding specific detection of target sequences. However, to date, this detection has been demonstrated only for naturally short, single-strand nucleic acids like microRNAs, for which the probe can be matched in size to the target.

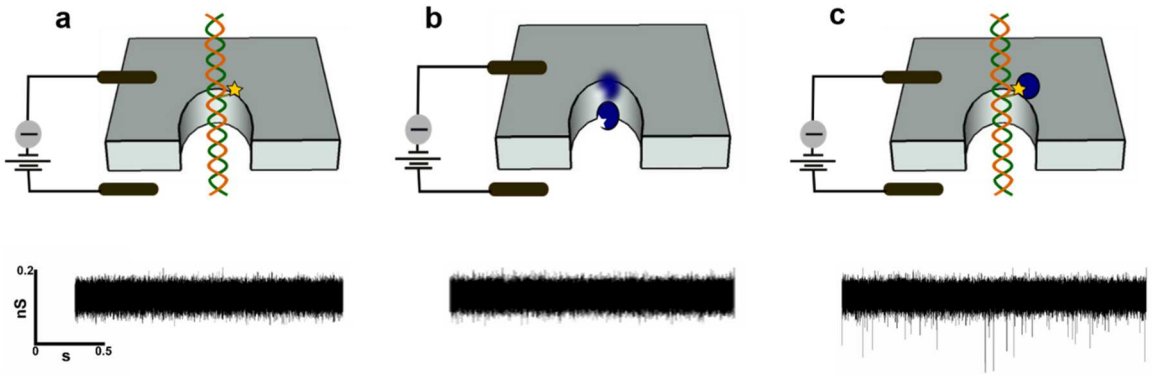


Figure 17: Selective solid-state nanopore assay. Diagrammatic representation (top) and example conductance trace (bottom) for the translocation of (a) individual biotinylated duplex nucleic acids, (b) monovalent streptavidin protein, and (c) duplex nucleic acids bound to monovalent streptavidin. Individual constituents transit through the nanopore rapidly and yield no events (a-b) while the nucleoprotein complex results in an event cascade due to steric hinderance. DNA translocations performed with a synthetic 60 bp construct and all traces recorded at 300 mV applied bias. Scale bar represents 0.2 nS and 0.5 s and applies to all traces.

In this report, we expand our protocols to enable the SS-nanopore assay to detect conserved sequence motifs within large single-strand genomes, including viral genomes. Our strategy, which involves annealing to a synthetic probe sequence and subsequent digestion of non-target regions, yields a short duplex construct capable of binding to MS and consequently being analyzed by our SS-nanopore assay. Like PCR, high specificity is achieved through Watson-Crick base pairing, but our approach provides potential advantages in both time and cost. We validate the technique using a model DNA genome (M13mp18 bacteriophage DNA) and then demonstrate its extension to RNA, the more common carrier of viral genetic information, by probing a conserved sequence in HIV. The ability of our approach to assess target sub-sequences within large DNAs and RNAs suggests the potential value of SS-nanopores not only for viral identification but also for other nucleic acid detection applications as well.

4.3 Materials and methods

4.3a Biomolecules

All synthetic oligonucleotides, including all synthetic probes and complementary sequences used to form control constructs were obtained commercially (Integrated DNA Technologies, Coralville, IA). The sequences of all probes are listed in Supplementary Table S1. All the oligonucleotides were resuspended in deionized water free of DNAse and RNAse (Invitrogen, Grand Island, NY) to a stock concentration of 200 μM and stored at -20ºC before use. single-stranded M13mp18 (New England Biolabs, Ipswich, MA). Synthetic duplex constructs were formed by annealing complementary oligonucleotides at a molar ratio of 1:1, heating to 95ºC for 10 minutes, and slowly cooling to room temperature.

HIV-1 BH10 non-infectious molecular clone (pBH10) was obtained from the NIH AIDS Reagent program (catalog #90) and subcloned into the SstI site of SP64 vector (Promega, Madison, WI) as a competent bacterial culture that was stored in glycerol stock at -80ºC prior to use [172]. 50 uL of bacterial stock was plated on an LB agar plate with a 50 ug/mL concentration of ampicillin and grown overnight in a bacterial incubator at 37ºC. Formed colonies were isolated, transferred to 5 mL of LB with 5 uL of 10mg/mL ampicillin, and then put in a shaker (135 rpm) at 37ºC for 12-16 hours until reaching an optical density (A_{600}) between 2-4. The bacterial culture was spun down at 5,000x *G* at 4ºC for 20 minutes to form a bacterial pellet which was isolated and used as a source for plasmid DNA using the Wizard SV Mini Prep Plasmid DNA Purification kit (Promega). 1 ug of the purified pbH10 plasmid was mixed with 1 U of NheI (New England Biolabs) and 3 μL of 10X NEBuffer 3.1 (New England Biolabs) in a 30 μL aliquot and incubated at 37ºC for 1 hr to cut the DNA downstream of the SP6 promoter and BH10 insert. After heat inactivating the NheI at 70ºC, a small aliquot of the product was

analyzed on a 1% agarose gel to ensure successful linearization. The linearized DNA was then processed with the HiScribe™ SP6 RNA Synthesis Kit (New England Biolabs) to synthesize full-length viral RNA product following the manufacturer's suggested protocol. 1 uL of SUPERase•In™ RNAse inhibitor (Invitrogen) was added prior to sample storage at -20^{o}C for stability. For all RNA processing steps, nuclease-free reagents were employed and work was carried out in an Airclean PCR workstation (AirClean Systems, Creedmore, NC) to reduce contamination.

4.3b DNA sequence isolation protocol

Annealing was carried out by mixing synthetic probes and M13mp18 DNA at a molar ratio of 13:1 in water, heating to 95^{o}C for 10 minutes, and then slowly cooling to room temperature. Digestion of unconjugated regions of M13mp18 and the excess probes was performed by adding 20 U MBN (New England Biolabs) for every 2.5 μg of M13mp18, adding 10X MBN Reaction Buffer (New England Biolabs) to reach a final concentration of 1X, and incubating the mixture at 30^{o}C for 15 mins. Endonuclease activity was quenched by adding an equal volume of phenol/chloroform/Isoamyl alcohal (25:24:1) (Acros Organics, Morris, NJ) followed by centrifugation at 17,000g in phase lock gel tubes (Quantabio, Beverly, MA). The process was repeated once using pure chloroform to remove any residual phenol. The aqueous component containing DNA isolate was retrieved and further purified using a Nucleotide Purification Kit (Qiagen, Germantown, MD). Obtained DNA constructs were redispersed in DNAse-/RNAse-free water at a final concentration of 150 nM and diluted as needed for concentration-dependent studies. The samples were loaded on a 3% agarose gel in 1X TBE buffer with Gel Red nucleic acid dye (Phenix Research Products, Candler, NC) and imaged by GelDoc™ system (BioRad, Hercules, CA) to confirm products.

4.3c RNA sequence isolation protocol

The multiple sequence alignment program Kalign (European Informatics Center) was used to align sequences of HIV-1B for identifying a candidate probe [173]. 4,300 full genome sequences of HIV-1B retrieved from Los Alamos National Laboratory (LANL) database were considered ranging in collection year and geographical area to maximize precision of the conserved sequence [174]. The target conserved motif was selected from candidate sequences with more than 90% alignment among all considered genomes. Annealing was performed using the same conditions as employed for M13mp18 but with a molar ratio of probe to RNA of 1:40. Magnesium chloride was added to the annealing mixture at a final concentration of 200 μM to enhance stability of duplex by screening the negatively charged phosphate backbones of the nucleic acids. After annealing, 0.2 U MBN was added for every 10 μg of RNA, 10X MBN Reaction Buffer was added to reach a final concentration of 1X, and the mixture was incubated for 15 mins at 30^{o}C. All other steps were followed as for M13mp18 above except that the use of the nucleotide removal kit for purification was replaced with an ethanol precipitation step to reduce contamination from spin columns and buffers. The obtained heteroduplex products were redispersed in DNAse-/RNAse-free water at a final concentration of 320 nM and diluted as needed for concentration-dependent measurements. Products were analyzed by gel electrophoresis as with M13mp18 as above but with the electrophoretic cells cleaned using a 0.1% SDS solution in 70% ethanol before gel casting.

4.3d Binding reaction between biotin tagged biomolecules and MS

Monovalent streptavidin (MS) was supplied by Howarth lab (Oxford University). For all binding reactions, duplex constructs were reacted with MS at a molar ratio of 1:4 in 1X PBS and incubated for 10 mins at room temperature [175]. The RNA-DNA hybrid construct was incubated with MS in a 1:40 molar ratio.

4.3e Solid State nanopores

4×4 mm silicon chips with a frame size and 20 μm silicon nitride window were obtained commercially (Norcada, Alberta, Canada). RNA measurements were performed on devices with 30 nm thick membranes while all other measurements shown used 20 nm thick membranes. In each membrane chip, a single nanopore 7.5-10 nm in diameter was fabricated using a scanning helium ion microscope (Orion PLUS, Carl Zeiss, Peabody, MA) technique described elsewhere and stored in 50% ethanol prior to use [176]. Directly before a measurement, chips were rinsed with ethanol and HPLC grade water, dried under filtered air flow, and cleaned with air plasma (30W) for 2 mins on each side to promote hydrophilicity. The silicon chip was then placed into a custom 3D printed flow cell and immediately introduced with measurement buffer (900 mM NaCl, 0.5X PBS) on both sides. For application of voltage and measurement of current, Ag/AgCl electrodes were immersed into the solution on each side and connected to a patch-clamp amplifier (Axopatch 200B, Molecular Devices, San Jose, CA). Pore diameters were verified through analysis of a linear current-voltage curve and comparison to a standard model [177]. Translocation measurements were carried out by replacing the measurement buffer in the *cis* chamber with analyte in measurement buffer. All current traces were collected at a bandwidth of 200 kHz with a 100 kHz four-pole Bessel filter and analyzed using custom software through which an additional 30 kHz low-pass filter was applied. Event rates were determined by

analyzing discrete 3.2 s blocks of data to identify events as defined by signals exceeding a threshold amplitude of 4.5σ and having a duration between 12.5 μs and 2 ms.

4.5 Results and discussion

4.5a Internal sequence isolation

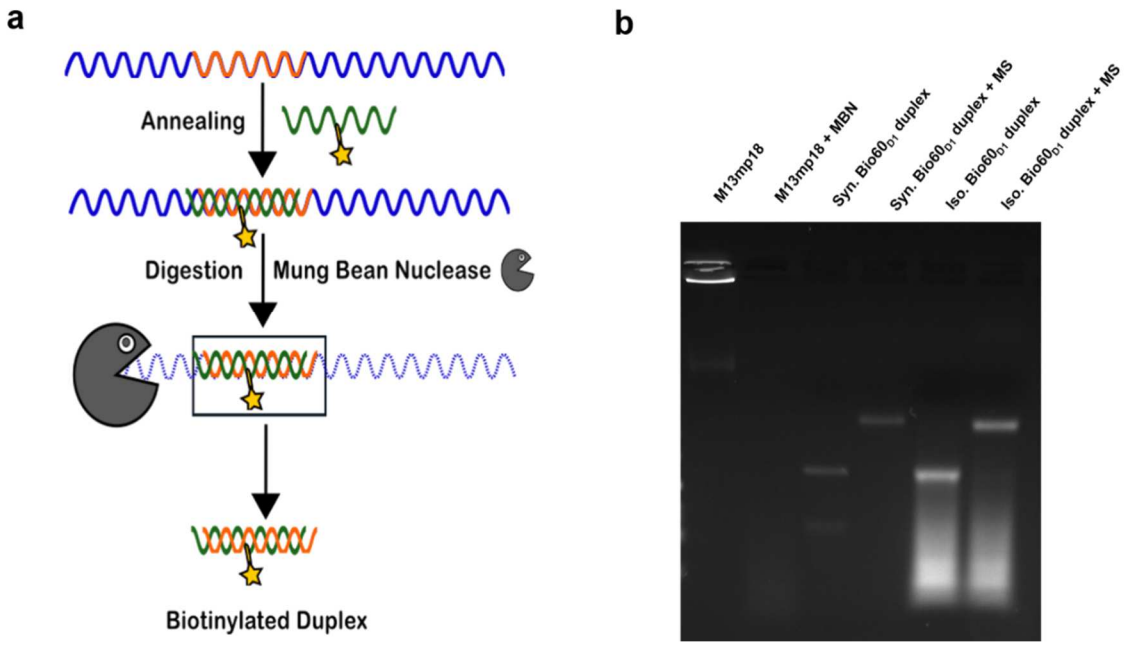


Figure 18: Visualization of the product from the digestion of hybridized molecular probe to genome. (a) Schematic representation of steps involved in the isolation of sub-sequences from a single-strand genome. A long genome (i) containing a target sequence motif (red) is annealed (ii) to a complimentary monobiotinylated DNA probe (green). The resulting construct is then subjected to digestion with a single-strand specific endonuclease (mung bean nuclease or MBN, iii) to yield a monobiotinylated duplex. (b) Agarose gel electrophoresis analysis. M13mp18 DNA alone (lane 1) is fully digested by MBN (lane 2). When annealed to a monobiotinylated 60 nt DNA probe, the same treatment yields a 60 bp duplex product (lane 5) that shifts upon binding to monovalent streptavidin (lane 6), identical to a synthetic monobiotinylated 60 bp duplex construct (lanes 3-4).

Our general approach to produce biotinylated duplex constructs from single-strand genomes is shown schematically in **Figure 18a**. First, a short target region is identified within the genome (**Figure 18a** *(i)*, red) and used to design a complementary probe construct (**Figure 18a** *(i)*, green)

consisting of a synthetic DNA oligonucleotide with a single biotinylated base. This probe is then mixed with the genomic material and subjected to a single thermal cycle to promote annealing, resulting in constructs that are locally duplex at the target site but single-strand elsewhere (**Figure 18a** *(ii)*). Finally, the constructs are incubated with mung bean nuclease (MBN), an endonuclease that catalyzes the degradation of internal phosphodiester bonds specifically in single-strand nucleic acids [178]. This treatment digests un-annealed genomic regions (**Figure 18a** *(iii)*) but leaves the target sequence intact, yielding duplex fragments matched in length to the synthetic probe and retaining the biotin tag for downstream analysis (**Figure 18a** *(iv)*). Note that positioning of the biotin near the center of the DNA probe reduces the chance of inadvertent MBN digestion due to thermally-driven structural fluctuations (or "breathing") of the duplex product at its ends.

We established our protocol initially using as a template a model single-strand DNA genome derived from the bacteriophage M13mp18 (7249 nt in length). For a first demonstration, we employed a 60 nt synthetic biotinylated probe designed with complementarity to an internal motif of M13mp18 (see **Materials and methods**) and determined assay success through gel analysis. M13mp18 alone yielded a well-defined band in the gel (**Figure 18b**, lane 1). Upon incubation with MBN (**Figure 18b**, lane 2), we observed a total loss of features in the lane, confirming the ability of the nuclease to digest the single-strand genomic DNA completely. Repeating the same procedure after annealing the biotinylated 60 nt probe to M13mp18 (**Figure 18b**, lane 5), we instead observed a distinct band that ran to the same position as a synthetic 60 bp construct (**Figure 18b**, lane 3), showing that the 60 bp construct is successfully isolated. Some low-molecular weight material was also seen following isolation (**Figure 18b**, lane 5 bottom), which we attribute to secondary structure elements in the DNA that would be slower to

digest due to their local duplex nature. Subsequent binding of the isolate to MS yielded a shift (**Figure 18b**, lane 6) identical to that of the synthetic construct (**Figure 18b**, lane 4), verifying the retention of an active biotin following nuclease treatment. Note that no shift was observed in the low-molecular weight products (**Figure 18b**, lane 5 bottom), indicating that they were composed of remnant template fragments rather than excess biotinylated probe.

MBN incubation time was critical for successful target isolation. When insufficient time or enzyme concentration were used, single-strand overhangs remained at the ends of the products as indicated by a smeared band on gel (**Figure** , **Appendix C**). This could negatively impact downstream processes. Conversely, extended incubation times could induce a loss of observed product (**Figure** , **Appendix C**). Even though the specificity of MBN for single-strand DNA is ~30,000× more than for duplex, deterioration could be driven by breathing at the ends that could provide stochastic single-strand recognition points for MBN activity [179].

4.5b SS-nanopore detection of target DNA motifs

Having demonstrated successful isolation of a target sub-sequence from a large single-strand DNA genome, we next demonstrated the viability of the isolated constructs to be analyzed with our selective SS-nanopore assay. We first performed a series of voltage-dependent measurements of translocation event rates on both the 60 bp product of M13mp18 isolation and MS alone (**Figure 19a**, black and green). We found that neither produced significant signals up to at least 400 mV. However, when combined, the resulting nucleoprotein construct produced clear events, the rate of which increased exponentially with applied voltage (**Figure 19a**, blue). These contrasting signals were in agreement with our past measurements using fully synthetic constructs and demonstrated our ability to detect a target internal DNA motif using the assay with high selectivity [171, 180]. Indeed, the event rate associated with MS-bound 60 bp isolate

was more than an order of magnitude higher than the rates of either constituent alone under all conditions.

We next conducted a concentration-dependent analysis to investigate quantitation with our assay. Using a single applied voltage (300 mV), we first performed a series of measurements on the biotinylated 60 bp M13mp18 product bound to MS. Across a range of 1 to 100 nM, we observed a linear trend in the measured event rate with a slope of 0.020 ± 0.001 $nM^{-1}s^{-1}$ (**Figure 19b**, blue). This trend matched our previous results with the same assay and was in quantitative agreement with similar measurements performed on a synthetic biotinylated 60 bp DNA construct bound to MS, which yielded a slope of 0.022 ± 0.002 $nM^{-1}s^{-1}$ (Fig. 3b, blue) [180]. The close correspondence demonstrated that our isolation protocol yielded an intact duplex DNA construct capable of SS-nanopore assay detection and indicated that targets could be quantified directly from event rate. Considering the low noise floor (i.e. the number of events associated with control measurements), we found that concentrations as low as 1 nM could be resolved easily without further improvement. Given the low fluid volume used in our system (10 μL), this corresponds to a total of 10 fmoles of product or an equivalent amount of single-strand genomes in solution.

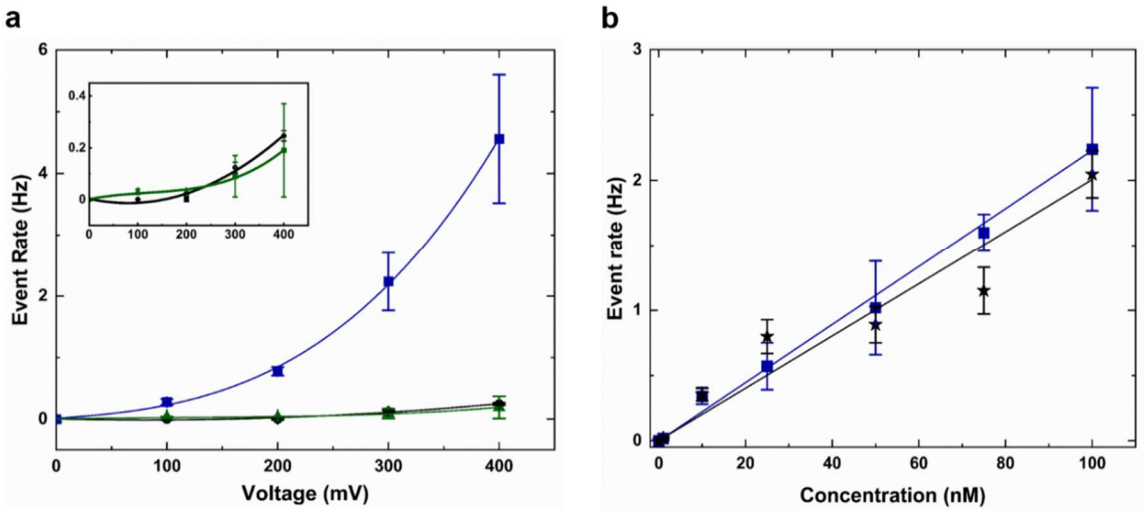


Figure 19: SS-nanopore analysis of 60 bp constructs isolated from M13mp18 DNA. (a) Measured translocation event rate vs applied voltage for MS alone (400nM, green), biotinylated 60 bp M13mp18 isolate (100 nM, black), and the combination following co-incubation (blue). Solid lines are exponential fits to the data. Inset: example current traces for each group recorded at 400 mV (colors matched). (b) Concentration dependence of event rate for biotinylated 60 bp M13mp18 isolate (blue) and a synthetic biotinylated 60 bp DNA construct (black) both bound to MS and at 300 mV applied voltage. Solid lines are linear fits to the data points.

4.5c Isolation and detection of multiple target motifs in a single genome

Because MBN is an endonuclease and does not require 3' or 5' ends to initiate, it should be possible for single-strand regions between multiple annealed probes along a single template molecule to be digested, thereby yielding multiple discrete products from one genome. To test this approach, we next repeated our isolation protocol with M13mp18 using three distinct synthetic probes. For easy differentiation on gel, three biotinylated probes with different lengths were utilized: the same 60 nt probe sequence employed above, as well as a 50 nt and a 75 nt probe with complementarity to disparate regions of the genomic sequence (see **Materials and methods**). Implementing an isolation protocol identical to that described above for M13mp18 but including all three probes during the anneal step, we subsequently observed three separate bands on gel (**Figure 20a**, lane 1), demonstrating the isolation of discrete products simultaneously. To verify the identities of these bands as the intended target constructs and to

confirm the retention of the biotin tags, two versions of the isolation procedure were performed: one in which biotinylated 60 nt probe and unlabeled 50 and 75 nt probes were used and one in which all three probe lengths were biotinylated. After mixing the resulting products with MS, we observed that only the center band shifted for the product of the first treatment, validating selectivity of the protocol. For the second treatment, all bands were found to shift, demonstrating consistent retention of the biotin moiety in all cases. In principle, this same process could be extended to arbitrary numbers of probes and sizes given sufficient sequence differentiability.

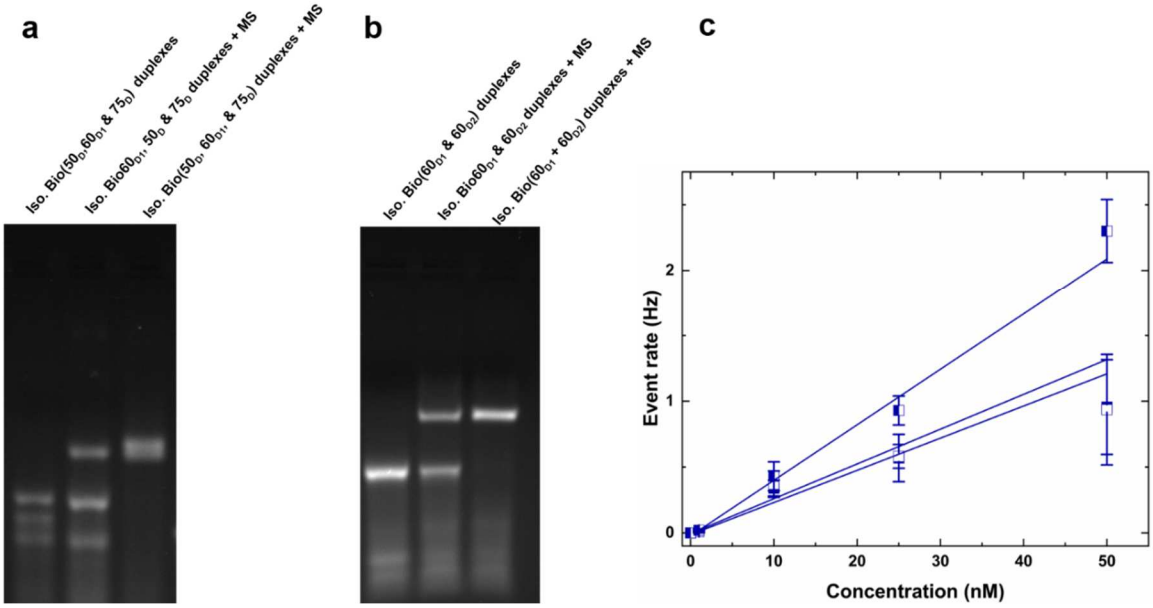


Figure 20: Isolate and analysis of multiple distinct products from a single-strand genome. (a) Gel analysis. Lane 1: isolated products from M13mp18 three DNA probes (50, 60 and 75 nt). Lane 2: electromobility shift assay in the presence of MS when only the 60 nt probe is biotinylated. Lane 3: electromobility shift assay in the presence of MS when all probes are biotinylated. (b) SS-nanopore assay signal amplification through the use of multiple probes. Concentration-dependent event rates for products of M13mp18 isolations using two independent biotinylated 60 nt probes separately (solid and open squares) and for products using both 60 nt probes in tandem (half solid and half white square) after binding with MS. All measurements performed at 300mV. Solid lines are linear fit to the data.

For the purpose of quantifying target genomes with our SS-nanopore assay, the simultaneous probing of multiple target sequences could provide a critical route to signal amplification without the need for PCR; when each genome provides multiple products, the overall sensitivity of detection should be improved concomitantly. We next demonstrated this concept empirically. Because translocation event rate is known to vary with construct length in our assay, the three probes used above for gel visualization could not be employed effectively here. Instead, we used two biotinylated synthetic probes of the same length (60 nt) that recognized different regions of M13mp18 (see **Materials and methods**) [180]. The biotinylated products of isolation for both probes were first confirmed on gel. When two probes were included in the anneal step, we retrieved a single band at the 60 bp position (**Figure 20b**, lane 1). While the brightness of the

observed band suggested qualitatively that more product was generated, the common size of both constructs made deconvolution challenging. To address this, we also performed the same isolation procedure with one biotinylated and one unmodified probe. In this case, binding to MS shifted only one of the two products (**Figure 20b**, lane 2), confirming that both target sequences were isolated independently. For completeness, when both probes were biotinylated, both were found to shift to a single position upon MS binding (**Figure 20b**, lane 3), as expected.

We next analyzed these products both separately and in tandem by SS-nanopore. When either 60 nt probe was used individually (**Figure 20c**, open, and solid squares), we observed a concentration-dependent event rate that was identical both to previous measurements (see **Figure 19b**) and to each other, exhibiting a slope of 0.026 ± 0.004, and 0.024 ± 0.003, respectively. However, when both probes were included concurrently, we recorded a dependence that was approximately double that of the single-probe measurements (**Figure 20c**, half solid and half open squares), yielding a slope of 0.042 ± 0.003. This confirmed that signal amplification could indeed be achieved through the use of multiple probes having complementarity to different regions of a single genome.

4.5d Detection of RNA sequence motifs

We finally extended our approaches toward pathogen detection by examining a viral genome. Because single-strand RNA viral genomes are significantly more abundant than single-strand DNA viruses, we chose to apply our protocols and subsequent nanopore analysis to a RNA genome. Here, we focused on HIV, a lentivirus (subcategory of retrovirus) consisting of an enveloped single-strand RNA. HIV is categorized into two types – HIV-1 and HIV-2 – with the former being a more virulent and major cause of infections globally [181, 182]. For this study, we specifically use as a model the viral genome of HIV-1B, the dominant subtype in the

Americas, Western Europe, Australia, and Asia [183]. While our DNA model M13mp18 had a stable, well-documented genome, HIV-1B is known to undergo mutation, making identification of a highly conserved target sequencing more challenging. We therefore first performed a sequence alignment of -4300 known variants. From this list, we identified a 60 nt motif in HIV-1B that was more than 90% conserved among all available sequences and designed a complementary biotinylated 60 nt DNA probe for use in our isolation protocol.

As an initial check, we performed an independent test of MBN nuclease on HIV-1B RNA alone because minimally the MBN digestion step of our procedure must degrade all of the template genome. Some available literature has reported that MBN activity is similar for DNA and RNA substrates, suggesting that our existing approach for isolation from DNA should hold [178, 184]. In contrast, Kole, *et al.* found that esters of ribonucleotides are cleaved by the enzyme 100 times faster than deoxyribonucleotides due to the presence of a 2'-hydroxyl group, suggesting that a significantly lower MBN exposure would be required for an RNA genome [185]. To test this directly, we performed a titration series of HIV-1B RNA against MBN (**Figure** , **Appendix C**) and found a sufficient level of degradation (**Figure 21a**, lane2) for the same incubation time as M13mp18 DNA but with a 400-fold lower concentration.

Taking this difference into account, we performed the full isolation protocol after annealing HIV-1B RNA with the biotinylated 60 nt probe and observed efficient recovery of the heteroduplex (RNA-DNA hybrid) product (**Figure 21a**, lane 5). We also observed an additional band below the 60 bp product that corresponded to unconjugated probe. Because of the strong reduction in MBN concentration used here, single-strand probes that were digested by the nuclease under DNA conditions were left largely intact. However, because single strand nucleic acids do not contribute to the SS-nanopore signal even when bound to MS, their presence was

not consequential to subsequent analyses [179]. As above, binding of the product with MS yielded an electromobility shift in gel analysis (**Figure 21a**, lane 6) that matched the behavior of a synthetic biotinylated DNA duplex (**Figure 21a**, lanes 3-4) and confirmed the presence of the biotin in the final constructs. To demonstrate the viability of our overall approach for RNA analysis, we finally conducted our SS-nanopore assay with the HIV-1B isolate. **Figure 21b** shows the result of concentration-dependent event rate measurements using the biotinylated heteroduplex product. We again find a linear relationship exhibiting a slope of 0.029±0.001. This result demonstrated extraordinary consistency with our measurements on M13mp18 above (c.f. **Figure 19b**), indicating that the sensitivity for RNA genomes should be identical to that of DNA genomes. As an additional control, we also measured

biotinylated 60 nt single-strand DNA probe bound to MS at a very high concentration (100 nM) and found that even that extreme amount yielded an event rate about an order of magnitude less than that of the duplex construct (**Figure 20b** inset). This confirmed that the undigested probe remaining after isolation from HIV-1B did not impact the measurement significantly.

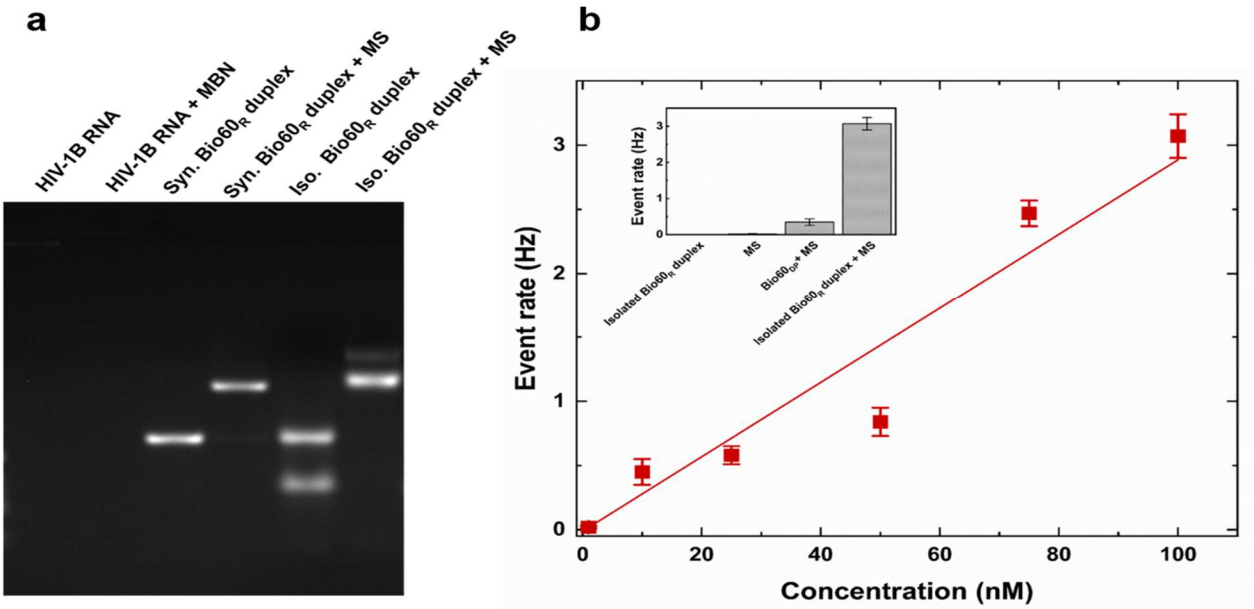


Figure 21: Sequence isolations and concentrations-dependent event rates for biotinylated HIV-1B RNA constructs. *(a)* Gel analysis of sequence isolation from HIV-1B RNA. Lane 1: HIV-1B RNA alone. Lane 2: HIV-1B RNA after MBN digestion. Lane 3: Synthetic biotinylated 60 bp DNA. Lane 4: Synthetic biotinylated 60 bp DNA bound to MS. Lane 5: biotinylated 60 bp RNA/DNA heteroduplex isolated from HIV-1B RNA. Lane 6: biotinylated 60 bp RNA/DNA heteroduplex isolated from HIV-1B RNA bound to MS. (b) Concentration-dependent event rate for monobiotinylated 60 bp RNA/DNA heteroduplex isolated from HIV-1B RNA bound to MS measured at 300 mV. Solid line is a linear fit to the data. Inset: event rates for 1: HIV-1B isolate alone (100 nM), 2: MS alone (4 μM), 3: 60 nt DNA probe bound to MS (100 nM), and 4: HIV-1B isolate bound MS. All measurements are carried out 300mV.

4.6 Conclusion

In this study, we have reported a new approach for the isolation of sequence motifs from single-strand genomes and their subsequent detection using a selective SS-nanopore assay. For isolation, a synthetic, biotinylated DNA oligonucleotide probe was designed with complementarity to a known sequence in the target genome. After annealing, the unpaired portions of the genomic template were digested using a single-strand specific endonuclease, resulting in duplex constructs matched in length to the probe. Because of the incorporated biotin moiety, these duplexes were amenable to analysis with a SS-nanopore assay in which the presence of a bound MS protein enables selective detection. We first demonstrated the approach

using as a model the single-strand bacteriophage DNA M13mp18, showing quantitative assessment of a single motif and extending to multiple independent sequence motifs in a common genome. We then applied the same concept to human pathogen detection by probing the HIV-1B genome. After adapting our protocols to work with RNA, we isolated a highly conserved sequence from the genome and observed similar results as with the DNA target, showing an ability to sense as little as 10 femtomoles of motif (and thus HIV-1B genomes) using only a single probe sequence.

Our method offers potential advantages over standard methods like PCR in that it is quantitative while being all-electronic (i.e. more cost effective than optical readouts) and could offer faster time-to-answer as it explicitly does not require amplification and the thermal cycling it entails. In addition, because we are able to probe RNA directly, our process does not require reverse transcription into cDNA prior to analysis. Ultimately, our results expand the capabilities of the approach and its associated protocols beyond the detection and quantification of short, size-matched targets like microRNA to include any single-strand targets [171]. This will enable numerous additional applications including pathogen screening of viral genomes, phylogeny and taxonomy assessments using ribosomal RNA, and transcriptional analysis of messenger RNA.

4.7 Acknowledgements

The following reagent was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 BH10 Non-infectious Molecular Clone (pBH10) from Dr. Beatrice Hahn and Dr. George Shaw (cat# 90). *Osama K. Zahid*

REFERENCES

1. Agamy, O., et al., *Mutations disrupting selenocysteine formation cause progressive cerebello-cerebral atrophy.* Am J Hum Genet, 2010. **87**(4): p. 538-44.
2. Taylor, E.W., et al., *Cellular Selenoprotein mRNA Tethering via Antisense Interactions with Ebola and HIV-1 mRNAs May Impact Host Selenium Biochemistry.* Curr Top Med Chem, 2016. **16**(13): p. 1530-5.
3. Berry, M.J., et al., *Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons.* EMBO J, 1993. **12**(8): p. 3315-22.
4. Donovan, J. and P.R. Copeland, *Evolutionary history of selenocysteine incorporation from the perspective of SECIS binding proteins.* BMC Evol Biol, 2009. **9**: p. 229.
5. Zhao, L., et al., *Molecular modeling and in vitro activity of an HIV-1-encoded glutathione peroxidase.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6356-61.
6. Burk, R.F. and K.E. Hill, *Selenoprotein P—Expression, functions, and roles in mammals.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2009. **1790**(11): p. 1441-1447.
7. Arthur, J.R., R.C. McKenzie, and G.J. Beckett, *Selenium in the immune system.* J Nutr, 2003. **133**(5 Suppl 1): p. 1457S-9S.
8. Berry, M.J., et al., *Selenocysteine insertion sequence element characterization and selenoprotein expression.* Methods Enzymol, 2002. **347**: p. 17-24.
9. Commans, S. and A. Böck, *Selenocysteine inserting tRNAs: an overview.* FEMS Microbiology Reviews, 1999. **23**(3): p. 335-351.
10. Ha, H.Y., et al., *From Selenium Absorption to Selenoprotein Degradation.* Biol Trace Elem Res, 2019. **192**(1): p. 26-37.
11. Hoffmann, P.R. and M.J. Berry, *Selenoprotein synthesis: a unique translational mechanism used by a diverse family of proteins.* Thyroid, 2005. **15**(8): p. 769-75.
12. Krol, A., *Evolutionarily different RNA motifs and RNA-protein complexes to achieve selenoprotein synthesis.* Biochimie, 2002. **84**(8): p. 765-74.
13. Seyfried, J. and U. Wullner, *Inhibition of thioredoxin reductase induces apoptosis in neuronal cell lines: role of glutathione and the MKK4/JNK pathway.* Biochem Biophys Res Commun, 2007. **359**(3): p. 759-64.
14. Guillin, O.M., et al., *Selenium, Selenoproteins and Viral Infection.* Nutrients, 2019. **11**(9).
15. Raman, A.V., et al., *Absence of selenoprotein P but not selenocysteine lyase results in severe neurological dysfunction.* Genes Brain Behav, 2012. **11**(5): p. 601-13.
16. Burk, R.F., et al., *Deletion of apolipoprotein E receptor-2 in mice lowers brain selenium and causes severe neurological dysfunction and death when a low-selenium diet is fed.* J Neurosci, 2007. **27**(23): p. 6207-11.
17. Byrns, C.N., et al., *Mice Lacking Selenoprotein P and Selenocysteine Lyase Exhibit Severe Neurological Dysfunction, Neurodegeneration, and Audiogenic Seizures.* Journal of Biological Chemistry, 2014. **289**(14): p. 9662-9674.
18. Hill, K.E., et al., *Neurological dysfunction occurs in mice with targeted deletion of the selenoprotein P gene.* J Nutr, 2004. **134**(1): p. 157-61.
19. Taylor, E.W., R.G. Nadimpalli, and C.S. Ramanathan, *Genomic structures of viral agents in relation to the biosynthesis of selenoproteins.* Biol Trace Elem Res, 1997. **56**(1): p. 63-91.
20. Zhang, W., et al., *Selenium-dependent glutathione peroxidase modules encoded by RNA viruses.* Biol Trace Elem Res, 1999. **70**(2): p. 97-116.

21. E.W., T. and R. J.A., *Zika-mediated antisense inhibition of selenoprotein synthesis may contribute to neurologic disorders and microcephaly by mimicking SePP1 knockout and the genetic disease PCCA.* Zika Open pre-print server, Bull. World Health Organ. , 2016. **E-pub:**: p. 13 July.
22. Dewitte, S.N., *Mortality Risk and Survival in the Aftermath of the Medieval Black Death.* PLoS ONE, 2014. **9**(5): p. e96513.
23. Nishiura, H., S.O. Brockmann, and M. Eichner, *Extracting key information from historical data to quantify the transmission dynamics of smallpox.* Theoretical Biology and Medical Modelling, 2008. **5**(1): p. 20.
24. Thompson, K.M., et al., *Trends in the Risk of U.S. Polio Outbreaks and Poliovirus Vaccine Availability for Response.* Public Health Reports, 2012. **127**(1): p. 23-37.
25. Morse, S.A., *Contagion and Chaos: Disease, Ecology, and National Security in the Era of Globalization.* Emerging Infectious Diseases, 2009. **15**(11): p. 1881b-1882.
26. Huremović, D., *Brief History of Pandemics (Pandemics Throughout History)*, in *Psychiatry of Pandemics*. 2019, Springer International Publishing. p. 7-35.
27. Chen, H.L. and R.B. Tang, *Why Zika virus infection has become a public health concern?* J Chin Med Assoc, 2016. **79**(4): p. 174-8.
28. Levin, P.A. and E.R. Angert, *Small but Mighty: Cell Size and Bacteria.* Cold Spring Harbor Perspectives in Biology, 2015. **7**(7): p. a019216.
29. Poltronieri, P., B. Sun, and M. Mallardo, *RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load.* Current Genomics, 2015. **16**(5): p. 327-335.
30. Kwun, H.J., et al., *Human DNA tumor viruses generate alternative reading frame proteins through repeat sequence recoding.* Proceedings of the National Academy of Sciences, 2014. **111**(41): p. E4342-E4349.
31. Olubajo, B. and E.W. Taylor, *A -1 frameshift in the HIV-1 env gene is enhanced by arginine deficiency via a hungry codon mechanism.* Mutat Res, 2005. **579**(1-2): p. 125-32.
32. Brierley, I., A.J. Jenner, and S.C. Inglis, *Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal.* Journal of Molecular Biology, 1992. **227**(2): p. 463-479.
33. Yu, C.-H., *Structural diversity of frameshifting signals: reprogramming the programmed.* 2011.
34. Hadi, K., et al., *A Mutagenesis Approach for the Study of the Structure-Function Relationship of Human Immunodeficiency Virus Type 1 (HIV-1) Vpr*, in *Genetic Manipulation of DNA and Protein - Examples from Current Research*. 2013, InTech.
35. Clarke, B.D., et al., *Functional non-coding RNAs derived from the flavivirus 3′ untranslated region.* Virus Res, 2015. **206**: p. 53-61.
36. Pijlman, G.P., et al., *A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity.* Cell Host Microbe, 2008. **4**(6): p. 579-91.
37. Schnettler, E., et al., *Noncoding flavivirus RNA displays RNA interference suppressor activity in insect and Mammalian cells.* J Virol, 2012. **86**(24): p. 13486-500.
38. Chapman, E.G., et al., *The Structural Basis of Pathogenic Subgenomic Flavivirus RNA (sfRNA) Production.* Science, 2014. **344**(6181): p. 307-310.
39. Taylor, E.W., et al., *A basis for new approaches to the chemotherapy of AIDS: novel genes in HIV-1 potentially encode selenoproteins expressed by ribosomal frameshifting and termination suppression.* J Med Chem, 1994. **37**(17): p. 2637-54.

40. Zhao, L., B. Olubajo, and E.W. Taylor, *Functional studies of an HIV-1 encoded glutathione peroxidase.* Biofactors, 2006. **27**(1-4): p. 93-107.
41. Cohen, I., et al., *Anti-apoptotic activity of the glutathione peroxidase homologue encoded by HIV-1.* Apoptosis, 2004. **9**(2): p. 181-92.
42. Uncini, A., N. Shahrizaila, and S. Kuwabara, *Zika virus infection and Guillain-Barré syndrome: a review focused on clinical and electrophysiological subtypes.* Journal of Neurology, Neurosurgery & Psychiatry, 2017. **88**(3): p. 266-271.
43. Faizan, M.I., et al., *Zika Virus-Induced Microcephaly and Its Possible Molecular Mechanism.* Intervirology, 2016. **59**(3): p. 152-158.
44. Ben-Zeev, B., et al., *Progressive cerebellocerebral atrophy: a new syndrome with microcephaly, mental retardation, and spastic quadriplegia.* J Med Genet, 2003. **40**(8): p. e96.
45. Faheem, M., et al., *Molecular genetics of human primary microcephaly: an overview.* BMC Med Genomics, 2015. **8 Suppl 1**: p. S4.
46. Leviton, A., et al., *Methodologic issues in epidemiologic studies of congenital microcephaly.* Early Hum Dev, 2002. **69**(1-2): p. 91-105.
47. Mochida, G.H. and C.A. Walsh, *Molecular genetics of human microcephaly.* Curr Opin Neurol, 2001. **14**(2): p. 151-6.
48. Barkovich, A.J., et al., *A developmental and genetic classification for malformations of cortical development: update 2012.* Brain, 2012. **135**(Pt 5): p. 1348-69.
49. Cardoso, B.R., et al., *Selenium, selenoproteins and neurodegenerative diseases.* Metallomics, 2015. **7**(8): p. 1213-28.
50. Guillin, O., et al., *Selenium, Selenoproteins and Viral Infection.* Nutrients, 2019. **11**(9): p. 2101.
51. Tang, H., et al., *Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth.* Cell Stem Cell, 2016. **18**(5): p. 587-90.
52. Calvet, G., et al., *Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study.* Lancet Infect Dis, 2016. **16**(6): p. 653-660.
53. Tabata, T., et al., *Zika Virus Targets Different Primary Human Placental Cells, Suggesting Two Routes for Vertical Transmission.* Cell Host Microbe, 2016. **20**(2): p. 155-66.
54. Lee, J. and O. Shin, *Advances in Zika Virus–Host Cell Interaction: Current Knowledge and Future Perspectives.* International Journal of Molecular Sciences, 2019. **20**(5): p. 1101.
55. Besnard, M., et al., *Evidence of perinatal transmission of Zika virus, French Polynesia, December 2013 and February 2014.* Euro Surveill, 2014. **19**(13).
56. Schwärzler, P., et al., *Late onset microcephaly: failure of prenatal diagnosis.* Ultrasound in Obstetrics and Gynecology, 2003. **22**(6): p. 640-642.
57. Arteel, G.E., et al., *Selenoprotein P.* Methods Enzymol, 2002. **347**: p. 121-5.
58. Burk, R.F. and K.E. Hill, *Selenoprotein P: an extracellular protein with unique physical characteristics and a role in selenium homeostasis.* Annu Rev Nutr, 2005. **25**: p. 215-35.
59. Byrns, C.N., et al., *Mice lacking selenoprotein P and selenocysteine lyase exhibit severe neurological dysfunction, neurodegeneration, and audiogenic seizures.* J Biol Chem, 2014. **289**(14): p. 9662-74.
60. Seeher, S., et al., *Impaired selenoprotein expression in brain triggers striatal neuronal loss leading to co-ordination defects in mice.* Biochem J, 2014. **462**(1): p. 67-75.

61. Zhang, X., et al., *Prioritized brain selenium retention and selenoprotein expression: Nutritional insights into Parkinson's disease.* Mech Ageing Dev, 2019. **180**: p. 89-96.
62. Ferenczy, M.W., et al., *Differentiation of Human Fetal Multipotential Neural Progenitor Cells to Astrocytes Reveals Susceptibility Factors for JC Virus.* Journal of Virology, 2013. **87**(11): p. 6221-6231.
63. Björkhem-Bergman, L., et al., *Prenatal expression of thioredoxin reductase 1 (TRXR1) and microsomal glutathione transferase 1 (MGST1) in humans.* FEBS Open Bio, 2014. **4**(1): p. 886-891.
64. Buckley, A. and E.A. Gould, *Detection of virus-specific antigen in the nuclei or nucleoli of cells infected with Zika or Langat virus.* J Gen Virol, 1988. **69 (Pt 8)**: p. 1913-20.
65. Chapman, E.G., et al., *The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production.* Science, 2014. **344**(6181): p. 307-10.
66. Finol, E. and E.E. Ooi, *Evolution of Subgenomic RNA Shapes Dengue Virus Adaptation and Epidemiological Fitness.* iScience, 2019. **16**: p. 94-105.
67. Funk, A., et al., *RNA structures required for production of subgenomic flavivirus RNA.* J Virol, 2010. **84**(21): p. 11407-17.
68. Rehmsmeier, M., et al., *Fast and effective prediction of microRNA/target duplexes.* RNA, 2004. **10**(10): p. 1507-17.
69. Busch, A., A.S. Richter, and R. Backofen, *IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions.* Bioinformatics, 2008. **24**(24): p. 2849-56.
70. Wright, P.R., et al., *CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W119-23.
71. Chittum, H.S., et al., *Multiple forms of selenoprotein P in rat plasma.* Arch Biochem Biophys, 1996. **325**(1): p. 124-8.
72. Read, R., et al., *Selenium and amino acid composition of selenoprotein P, the major selenoprotein in rat serum.* J Biol Chem, 1990. **265**(29): p. 17899-905.
73. Motchnik, P.A. and A.L. Tappel, *Rat plasma selenoprotein P properties and purification.* Biochim Biophys Acta, 1989. **993**(1): p. 27-35.
74. Burk, R.F. and K.E. Hill, *Selenoprotein P. A selenium-rich extracellular glycoprotein.* J Nutr, 1994. **124**(10): p. 1891-7.
75. Jacobson, G.A., et al., *Selenoprotein P analysis in human plasma: a discrepancy between HPLC fractionation of human plasma with heparin-affinity chromatography and SDS-PAGE with immunoblot analysis.* Biol Trace Elem Res, 2005. **107**(3): p. 213-20.
76. Burk, R.F., et al., *Response of rat selenoprotein P to selenium administration and fate of its selenium.* Am J Physiol, 1991. **261**(1 Pt 1): p. E26-30.
77. Thanbichler, M., *The function of SECIS RNA in translational control of gene expression in Escherichia coli.* The EMBO Journal, 2002. **21**(24): p. 6925-6934.
78. Uchil, P.D., A.V. Kumar, and V. Satchidanandam, *Nuclear localization of flavivirus RNA synthesis in infected cells.* J Virol, 2006. **80**(11): p. 5451-64.
79. Lopez-Denman, A. and J. Mackenzie, *The IMPORTance of the Nucleus during Flavivirus Replication.* Viruses, 2017. **9**(1): p. 14.
80. de Jesus, L.A., et al., *Nuclear assembly of UGA decoding complexes on selenoprotein mRNAs: a mechanism for eluding nonsense-mediated decay?* Mol Cell Biol, 2006. **26**(5): p. 1795-805.

81. Branco, V., et al., *Mitochondrial thioredoxin reductase inhibition, selenium status, and Nrf-2 activation are determinant factors modulating the toxicity of mercury compounds.* Free Radical Biology and Medicine, 2014. **73**: p. 95-105.
82. Fang, L.Q., et al., *The association between hantavirus infection and selenium deficiency in mainland China.* Viruses, 2015. **7**(1): p. 333-51.
83. Hill, K.E., et al., *Selenoprotein P Concentration in Plasma is an Index of Selenium Status in Selenium-Deficient and Selenium-Supplemented Chinese Subjects.* The Journal of Nutrition, 1996. **126**(1): p. 138-145.
84. Hill, K.E., et al., *The Selenium-rich C-terminal Domain of Mouse Selenoprotein P Is Necessary for the Supply of Selenium to Brain and Testis but Not for the Maintenance of Whole Body Selenium.* Journal of Biological Chemistry, 2007. **282**(15): p. 10972-10980.
85. Schweizer, U., et al., *Efficient selenium transfer from mother to offspring in selenoprotein-P-deficient mice enables dose-dependent rescue of phenotypes associated with selenium deficiency.* Biochem J, 2004. **378**(Pt 1): p. 21-6.
86. Wichmann, D., et al., *Hantaan Virus Infection Causes an Acute Neurological Disease That Is Fatal in Adult Laboratory Mice.* Journal of Virology, 2002. **76**(17): p. 8890-8899.
87. Fang, L.-Q., et al., *The Association between Hantavirus Infection and Selenium Deficiency in Mainland China.* Viruses, 2015. **7**(1): p. 333-351.
88. Hughes, D., et al., *Expression of Selenoprotein Genes and Association with Selenium Status in Colorectal Adenoma and Colorectal Cancer.* Nutrients, 2018. **10**(11): p. 1812.
89. Lothrop, A.P., et al., *Why Is Mammalian Thioredoxin Reductase 1 So Dependent upon the Use of Selenium?* Biochemistry, 2014. **53**(3): p. 554-565.
90. Ganther, H.E., *Selenium metabolism, selenoproteins and mechanisms of cancer prevention: complexities with thioredoxin reductase.* Carcinogenesis, 1999. **20**(9): p. 1657-1666.
91. Branco, V., et al., *Mercury and selenium interaction in vivo: Effects on thioredoxin reductase and glutathione peroxidase.* Free Radical Biology and Medicine, 2012. **52**(4): p. 781-793.
92. Nalvarte, I., et al., *Overexpression of Enzymatically Active Human Cytosolic and Mitochondrial Thioredoxin Reductase in HEK-293 Cells.* Journal of Biological Chemistry, 2004. **279**(52): p. 54510-54517.
93. Sroka, J., et al., *Overexpression of thioredoxin reductase 1 inhibits migration of HEK-293 cells.* Biology of the Cell, 2007. **99**(12): p. 677-687.
94. Rundlof, A.K. and E.S. Arner, *Regulation of the mammalian selenoprotein thioredoxin reductase 1 in relation to cellular phenotype, growth, and signaling events.* Antioxid Redox Signal, 2004. **6**(1): p. 41-52.
95. Tobe, R., et al., *Thioredoxin reductase 1 deficiency enhances selenite toxicity in cancer cells via a thioredoxin-independent mechanism.* Biochemical Journal, 2012. **445**(3): p. 423-430.
96. Poerschke, R.L. and P.J. Moos, *Thioredoxin reductase 1 knockdown enhances selenazolidine cytotoxicity in human lung cancer cells via mitochondrial dysfunction.* Biochemical Pharmacology, 2011. **81**(2): p. 211-221.
97. Vasireddi, M., et al., *A novel antiviral inhibits Zika virus infection while increasing intracellular glutathione biosynthesis in distinct cell culture models.* Antiviral Research, 2019. **161**: p. 46-52.

98. Kalantari, P., et al., *Thioredoxin Reductase-1 Negatively Regulates HIV-1 Transactivating Protein Tat-dependent Transcription in Human Macrophages.* Journal of Biological Chemistry, 2008. **283**(48): p. 33183-33190.
99. Richard, A.S., et al., *AXL-dependent infection of human fetal endothelial cells distinguishes Zika virus from other pathogenic flaviviruses.* Proceedings of the National Academy of Sciences, 2017. **114**(8): p. 2024-2029.
100. Strange, D.P., et al., *Axl Promotes Zika Virus Entry and Modulates the Antiviral State of Human Sertoli Cells.* mBio, 2019. **10**(4).
101. Duan, D., et al., *Targeting Thioredoxin Reductase by Parthenolide Contributes to Inducing Apoptosis of HeLa Cells.* Journal of Biological Chemistry, 2016. **291**(19): p. 10021-10031.
102. Liu, R., et al., *Xanthatin Promotes Apoptosis via Inhibiting Thioredoxin Reductase and Eliciting Oxidative Stress.* Molecular Pharmaceutics, 2018. **15**(8): p. 3285-3296.
103. Zou, P., et al., *EF24 induces ROS-mediated apoptosis via targeting thioredoxin reductase 1 in gastric cancer cells.* Oncotarget, 2016. **7**(14): p. 18050-18064.
104. Angstwurm, M.W., et al., *Selenium in Intensive Care (SIC): results of a prospective randomized, placebo-controlled, multiple-center study in patients with severe systemic inflammatory response syndrome, sepsis, and septic shock.* Crit Care Med, 2007. **35**(1): p. 118-26.
105. Baum, M.K., et al., *High risk of HIV-related mortality is associated with selenium deficiency.* J Acquir Immune Defic Syndr Hum Retrovirol, 1997. **15**(5): p. 370-4.
106. Beck, M.A., et al., *Selenium deficiency increases the pathology of an influenza virus infection.* FASEB J, 2001. **15**(8): p. 1481-3.
107. Campa, A., G. Shor-Posner, and M. Baum, *Selenium status and the human immunodeficiency virus.* J Am Diet Assoc, 2000. **100**(4): p. 418.
108. Murai, K., et al., *Induction of Selenoprotein P mRNA during Hepatitis C Virus Infection Inhibits RIG-I-Mediated Antiviral Immunity.* Cell Host & Microbe, 2019. **25**(4): p. 588-601.e7.
109. Khromykh, A.A. and E.G. Westaway, *Subgenomic replicons of the flavivirus Kunjin: construction and applications.* J Virol, 1997. **71**(2): p. 1497-505.
110. Gonzalez-Flores, J.N., et al., *The molecular biology of selenocysteine.* BioMolecular Concepts, 2013. **4**(4): p. 349-365.
111. Faye, O., et al., *Quantitative real-time PCR detection of Zika virus and evaluation with field-caught Mosquitoes.* Virology Journal, 2013. **10**(1): p. 311.
112. Buchan, B.W., et al., *Distribution of SARS-CoV-2 PCR Cycle Threshold Values Provide Practical Insight Into Overall and Target-Specific Sensitivity Among Symptomatic Patients.* American Journal of Clinical Pathology, 2020. **154**(4): p. 479-485.
113. Araujo, R.V., et al., *One-step RT-qPCR assay for ZIKV RNA detection in Aedes aegypti samples: a protocol to study infection and gene expression during ZIKV infection.* Parasites & Vectors, 2020. **13**(1).
114. Baum, M.K., et al., *Effect of micronutrient supplementation on disease progression in asymptomatic, antiretroviral-naive, HIV-infected adults in Botswana: a randomized clinical trial.* JAMA, 2013. **310**(20): p. 2154-63.
115. Jiamton, S., et al., *A randomized trial of the impact of multiple micronutrient supplementation on mortality among HIV-infected individuals living in Bangkok.* AIDS, 2003. **17**(17): p. 2461-9.

116. Beck, M.A., et al., *Increased virulence of a human enterovirus (coxsackievirus B3) in selenium-deficient mice.* J Infect Dis, 1994. **170**(2): p. 351-7.
117. Zhong, L., E.S. Arner, and A. Holmgren, *Structure and mechanism of mammalian thioredoxin reductase: the active site is a redox-active selenolthiol/selenenylsulfide formed from the conserved cysteine-selenocysteine sequence.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5854-9.
118. Gonzalez-Flores, J.N., et al., *The molecular biology of selenocysteine.* Biomol Concepts, 2013. **4**(4): p. 349-65.
119. Taylor, E.W., et al., *Nutrition, HIV, and drug abuse: the molecular basis of a unique role for selenium.* J Acquir Immune Defic Syndr, 2000. **25 Suppl 1**: p. S53-61.
120. Feng, Y.X., et al., *Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich sequence immediately downstream of the gag termination codon followed by an RNA pseudoknot.* J Virol, 1992. **66**(8): p. 5127-32.
121. Su, G., W. Min, and E.W. Taylor, *An HIV-1 encoded peptide mimics the DNA binding loop of NF-kappaB and binds thioredoxin with high affinity.* Mutat Res, 2005. **579**(1-2): p. 133-48.
122. Abramoff, M.D., P.J. Magalhaes, and S.J. Ram, *Image Processing with ImageJ.* Biophotonics International, 2004. **11**(7): p. 36-42.
123. Chittum, H.S., et al., *Rabbit beta-globin is extended beyond its UGA stop codon by multiple suppressions and translational reading gaps.* Biochemistry, 1998. **37**(31): p. 10866-70.
124. Zeng, Y., R. Yi, and B.R. Cullen, *MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms.* Proc Natl Acad Sci U S A, 2003. **100**(17): p. 9779-84.
125. Dworkin, B.M., et al., *Selenium deficiency in the acquired immunodeficiency syndrome.* JPEN J Parenter Enteral Nutr, 1986. **10**(4): p. 405-7.
126. Dworkin, B.M., et al., *Abnormalities of blood selenium and glutathione peroxidase activity in patients with acquired immunodeficiency syndrome and aids-related complex.* Biol Trace Elem Res, 1988. **15**: p. 167-77.
127. Campa, A., et al., *Mortality risk in selenium-deficient HIV-positive children.* J Acquir Immune Defic Syndr Hum Retrovirol, 1999. **20**(5): p. 508-13.
128. Constans, J., et al., *Serum selenium predicts outcome in HIV infection.* J Acquir Immune Defic Syndr Hum Retrovirol, 1995. **10**(3): p. 392.
129. Kamwesiga, J., et al., *Effect of selenium supplementation on CD4+ T-cell recovery, viral suppression and morbidity of HIV-infected patients in Rwanda: a randomized controlled trial.* AIDS, 2015. **29**(9): p. 1045-52.
130. L., H. *Selenium and RNA viruses interactions: Potential implications for SARS-Cov-2 infection (Covid-19)*. 2020. DOI: 10.31219/osf.io/vaqz6.
131. Steinbrenner, H., et al., *Dietary Selenium in Adjuvant Therapy of Viral and Bacterial Infections.* Adv Nutr, 2015. **6**(1): p. 73-82.
132. Zhang, J., et al., *Association between regional selenium status and reported outcome of COVID-19 cases in China.* Am J Clin Nutr, 2020.
133. M, O., *History of Virology*. Encyclopedia of Microbiology. 2014. 608-612.
134. Woolhouse, M., et al., *Human viruses: discovery and emergence.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2012. **367**(1604): p. 2864-2871.

135. Rosenberg, R., *Detecting the emergence of novel, zoonotic viruses pathogenic to humans.* Cellular and Molecular Life Sciences, 2015. **72**(6): p. 1115-1125.
136. Woolhouse, M. and E. Gaunt, *Ecological Origins of Novel Human Pathogens.* Critical Reviews in Microbiology, 2007. **33**(4): p. 231-242.
137. Sudhan, S.S., Sharma, P, *Human Viruses: Emergence and Evolution.* Emerg Infect Dis, 2020: p. 53-68.
138. Kiselev, D., Matsvay, A., Abramov, I., Dedkov, V., Shipulin, G., & Khafizov, K, *Current Trends in Diagnostics of Viral Infections of Unknown Etiology.* Viruses, 2020. **12**(2): p. 211.
139. Messina, F.A., et al., *Coronavirus Disease 2019 (COVID-19) in a Patient with Disseminated Histoplasmosis and HIV-A Case Report from Argentina and Literature Review.* J Fungi (Basel), 2020. **6**(4).
140. Mackay, I.M., *Real-time PCR in virology.* Nucleic Acids Research, 2002. **30**(6): p. 1292-1305.
141. Green, T.A., C.M. Black, and R.E. Johnson, *Evaluation of Bias in Diagnostic-Test Sensitivity and Specificity Estimates Computed by Discrepant Analysis.* Journal of Clinical Microbiology, 1998. **36**(2): p. 375-381.
142. Yang, S. and R.E. Rothman, *PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings.* The Lancet Infectious Diseases, 2004. **4**(6): p. 337-348.
143. Storch, G.A., *Diagnostic Virology.* Clinical Infectious Diseases, 2000. **31**(3): p. 739-751.
144. Dwenger, A., *Radioimmunoassay: an overview.* J Clin Chem Clin Biochem, 1984. **22**(12): p. 883-94.
145. Sun, Q.Z., G. Dou, W. , *Blue Silica Nanoparticle-Based Colorimetric Immunoassay for Detection of Salmonella Pullorum.* Analytical Methods, 2015. **7**: p. 8647-8654.
146. Hoofnagle, A.N.W., M. H. , *The Fundamental Flaws of Immunoassays and Potential Solutions Using Tandem Mass Spectrometry.* J. Immunology Methods, 2009. **347**: p. 3-11.
147. Dekker, C., *Solid-State Nanopores.* Nature Nanotechnology, 2007. **2**: p. 209-215.
148. Wanunu, M., *Nanopores: A Journey towards DNA Sequencing.* Physical Life Rev, 2012. **9**: p. 125-158.
149. Li, J., et al., *DNA molecules and configurations in a solid-state nanopore microscope.* Nature Materials, 2003. **2**(9): p. 611-615.
150. Storm, A.J., et al., *Translocation of double-strand DNA through a silicon oxide nanopore.* Physical Review E, 2005. **71**(5).
151. Skinner, G.M.v.d.H., M. Broekmans, O. Dekker, C. Dekker, N. H., *Distinguishing Single and Double Stranded Nucleic Acid Molecules Using Solid-State Nanopores.* Nano Letters, 2009. **9**: p. 2953-2960.
152. Fologea, D., et al., *Electrical characterization of protein molecules by a solid-state nanopore.* Applied Physics Letters, 2007. **91**(5): p. 053901.
153. Plesa, C., et al., *Fast Translocation of Proteins through Solid State Nanopores.* Nano Letters, 2013. **13**(2): p. 658-663.
154. Larkin, J., et al., *High-Bandwidth Protein Analysis Using Solid-State Nanopores.* Biophysical Journal, 2014. **106**(3): p. 696-704.

155. Anderson, W., et al., *Observations of Tunable Resistive Pulse Sensing for Exosome Analysis: Improving System Sensitivity and Stability.* Langmuir, 2015. **31**(23): p. 6577-6587.
156. Prabhu, A.S.J., T. Z. N. Freedman, K. J. Mulero, R. Dutta, P. Kim, M. J., *Chemically Modified Solid State Nanopores for High Throughput Nanoparticle Separation.* J. Physical-Condens. Matter, 2010. **22**.
157. Bacri, L.O., A. G. Schiedt, B. Patriarche, G. Bourhis, E. Gierak, J. Pelta, J. and L. Auvray, *Dynamics of Colloids in Single Solid-State Nanopores.* J. Physical Chemistry B., 2011. **115**: p. 2890-2898.
158. Arjmandi, N., W. Van Roy, and L. Lagae, *Measuring Mass of Nanoparticles and Viruses in Liquids with Nanometer-Scale Pores.* Analytical Chemistry, 2014. **86**(10): p. 4637-4641.
159. Mcmullen, A., et al., *Stiff filamentous virus translocations through solid-state nanopores.* Nature Communications, 2014. **5**(1).
160. Yang, L. and T. Yamamoto, *Quantification of Virus Particles Using Nanopore-Based Resistive-Pulse Sensing Techniques.* Frontiers in Microbiology, 2016. **7**.
161. Uram, J.D.K., K. Hunt, A. J. Mayer, M. , *Submicrometer Pore-Based Characterization and Quantification of Antibody-Virus Interactions.* Small Weinh. Bergstr. Ger, 2006. **2**: p. 967-972.
162. Zhou, K., et al., *Characterization of Hepatitis B Virus Capsids by Resistive-Pulse Sensing.* Journal of the American Chemical Society, 2011. **133**(6): p. 1618-1621.
163. Darvish, A., et al., *Mechanical characterization of HIV-1 with a solid-state nanopore sensor.* ELECTROPHORESIS, 2019. **40**(5): p. 776-783.
164. Arima, A., et al., *Selective detections of single-viruses using solid-state nanopores.* Scientific Reports, 2018. **8**(1).
165. Arima, A., et al., *Identifying Single Viruses Using Biorecognition Solid-State Nanopores.* Journal of the American Chemical Society, 2018. **140**(48): p. 16834-16841.
166. Nouri, R., et al., *Sequence-Specific Recognition of HIV-1 DNA with Solid-State CRISPR-Cas12a-Assisted Nanopores (SCAN).* ACS Sensors, 2020. **5**(5): p. 1273-1280.
167. Niedzwiecki, D.J., et al., *Sampling a Biomarker of the Human Immunodeficiency Virus across a Synthetic Nanopore.* ACS Nano, 2013. **7**(4): p. 3341-3350.
168. Carlsen, A.T., et al., *Selective Detection and Quantification of Modified DNA with Solid-State Nanopores.* Nano Letters, 2014. **14**(10): p. 5488-5492.
169. Howarth, M., et al., *A monovalent streptavidin with a single femtomolar biotin binding site.* Nature Methods, 2006. **3**(4): p. 267-273.
170. Fairhead, M.K., D. Lowe, E. D. Howarth, M., *Plug-and-Play Pairing via Defined Divalent Streptavidins.* J. Molecular Biology, 2014. **426**: p. 199-214.
171. Zahid, O.K., et al., *Sequence-Specific Recognition of MicroRNAs and Other Short Nucleic Acids with Solid-State Nanopores.* Nano Letters, 2016. **16**(3): p. 2033-2039.
172. Hahn, B.H.S., G. M. Arya, S. K. Popovic, M. Gallo, R. C. Wong-Staal, F., *Molecular Cloning and Characterization of the HTLV-III Virus Associated with AIDS.* Nature Communications, 1984. **312**: p. 166-169.
173. Lassmann, T., O. Frings, and E.L.L. Sonnhammer, *Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features.* Nucleic Acids Research, 2009. **37**(3): p. 858-865.
174. *Los Alamos National Laboratory. HIV Sequence Database* https://www.hiv.lanl.gov.

175. Yang, J.F., D. C.; Stern, L. A.; Sanford, C. A.; Huang, J.; Ren, Z.; Qin, L.-C.; Hall, and A. R., *Rapid and Precise Scanning Helium Ion Microscope Milling of Solid-State nanopores for Biomolecule Detection.* Nanotechnology, 2011. **22**: p. 285-310.
176. Yang, J.F., D. C. Stern, L. A. Sanford, C. A. Huang, J. Ren, Z. Qin, L.-C. Hall, and A. R., *Rapid and Precise Scanning Helium Ion Microscope Milling of Solid-Stae Nanopores for Biomolecule Detection.* Nanotechnology, 2011. **22**(28531).
177. Wanunu, M., et al., *Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors.* Nature Nanotechnology, 2010. **5**(11): p. 807-814.
178. Desai, N.A. and V. Shankar, *Single-strand-specific nucleases.* FEMS Microbiology Reviews, 2003. **26**(5): p. 457-491.
179. Ardelt, W. and M. Laskowski, *Mung bean nuclease I. IV. An improved method of preparation.* Biochemical and Biophysical Research Communications, 1971. **44**(5): p. 1205-1211.
180. Zahid, O.K., et al., *Quantifying mammalian genomic DNA hydroxymethylcytosine content using solid-state nanopores.* Scientific Reports, 2016. **6**(1): p. 29565.
181. Fauci, A., *The human immunodeficiency virus: infectivity and mechanisms of pathogenesis.* Science, 1988. **239**(4840): p. 617-622.
182. Nyamweya, S.H., A. Jaye, A. Rowland-Jones, S. Flanagan, K. L. Macallan, D.C., *Comparing HIV-1 and HIV-2 Infection: Lessons for Viral Immunopathogenesis.* Review Medical Virology, 2013. **23**: p. 221-240.
183. Junqueira, D.M. and S.E.D.M. Almeida, *HIV-1 subtype B: Traces of a pandemic.* Virology, 2016. **495**: p. 173-184.
184. Laskowski, M., *Purification and Properties of the Mung Bean Nuclease.* Methods Enzymol, 1980. **65**: p. 263-276.
185. Kole, R.S., H. Szemplińska, H. Shugar, D. , *Mung Bean Nuclease: Mode of Action and Specificity vs Synthetic Esters of 3'-Nucleotides.* Nucleic Acids Res, 1974. **1**: p. 699-706.

APPENDIX A:

SUPPLEMENTAL FIGURES AND TABLES FOR CHAPTER II

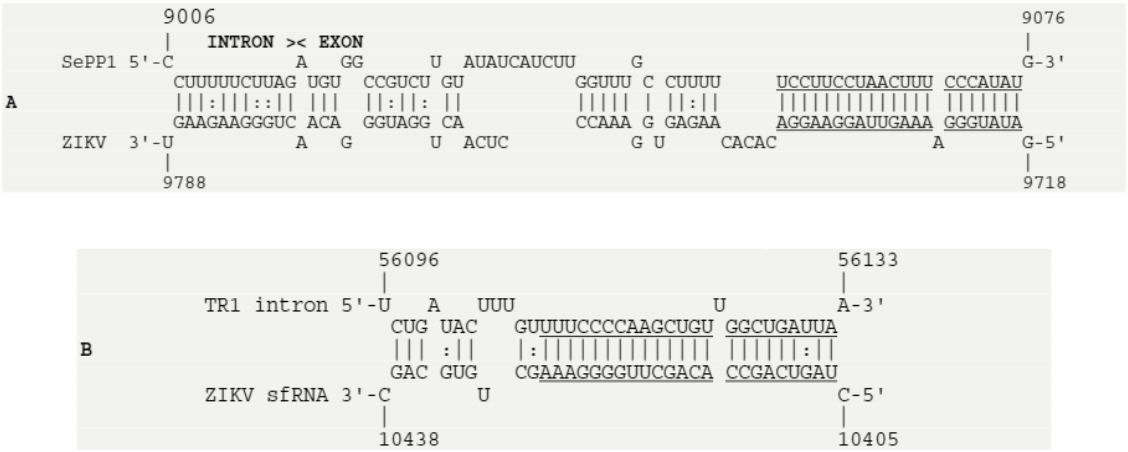
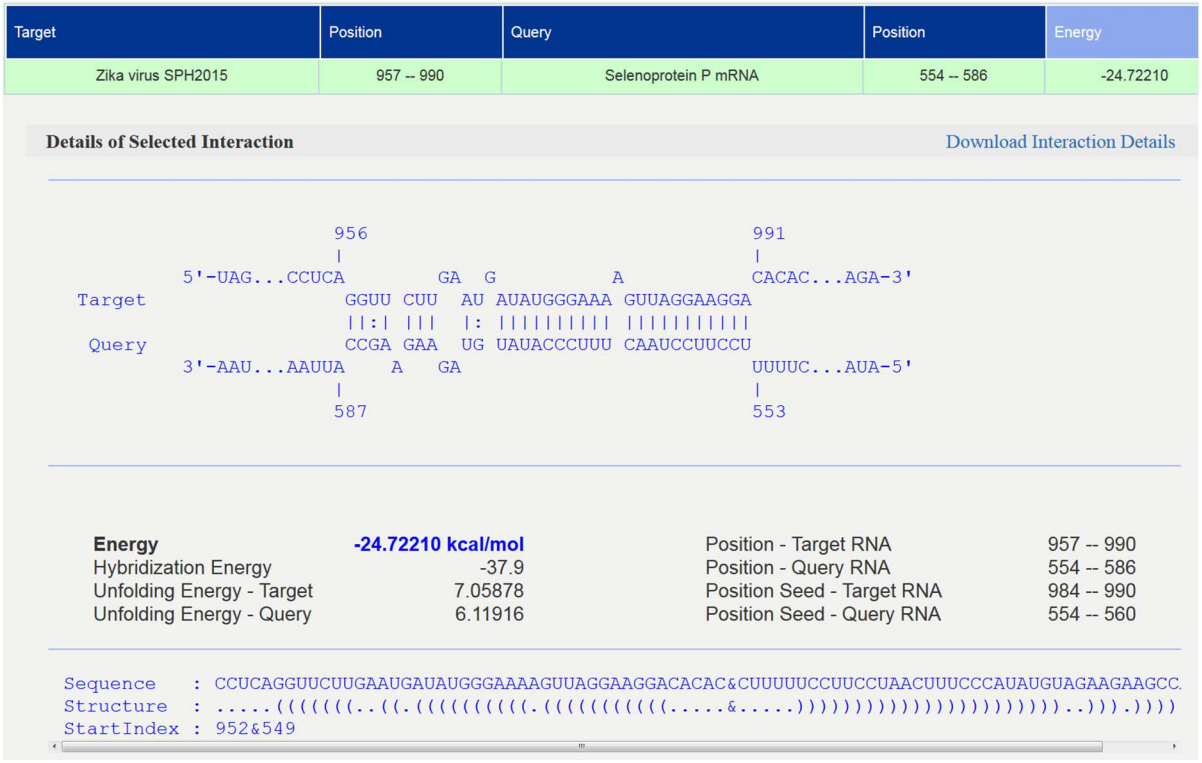


Figure A1: IntaRNA-predicted RNA-RNA interactions between Zika virus mRNA and regions of unspliced selenoprotein pre-mRNAs, confirming matches shown in Figure 5. Panal A shows the predicted interaction between a 3' region of the ZIKV mRNA (polymerase coding region) and a region of SePP1 pre-mRNA, spanning an intron/exon boundary in SePP1. This is essentially the same structure predicted by RNAHybrid shown in **Figure 5**, but slightly truncated at the SePP1 3' end. The SePP1 intron/exon boundary is indicted by the >< symbol above the sequence. For this analysis, a 500-base fragment of ZIKV (9,501-10,000) was scanned vs. a 2,000-base fragment (8,001-10,000) of the SePP1 genomic sequence. The computed hybridization energy for the structure shown is -59.6 kcal/mole, with a net energy of -25.7 kcal/mole, after subtraction of unfolding energies of 10.0 and 23.9 kcal/mole for the respective SePP1 and ZIKV RNA regions. The underlined region indicates the core of the antisense pairing of these two RNAs that was first identified by BLAST search as a 21/22 identity plus/minus sequence match, which is substantially extended by IntaRNA into the intron at left. In Panel B, the predicted interaction between a region of the ZIKV non-coding "subgenomic flavivirus RNA" (sfRNA) and an intronic region of the human thioredoxin reductase 1 (TR1) pre-mRNA. For this analysis, the ZIKV ~300-base 3'UTR region was scanned vs. a 2000-base fragment (55001-57000) of the TXNRD1 genomic sequence. The computed hybridization energy for the structure shown is -48.0 kcal/mole, with a net energy of -24.8 kcal/mole, after subtraction of unfolding energies of 7.6 and 15.6 kcal/mole for the respective TR1 and ZIKV RNA regions. The underlined region indicates the core of the antisense pairing of these two RNAs that was first identified by BLAST search as a 22/24 identity plus/minus sequence match, which is only slightly extended by IntaRNA.



Target	Position	Query	Position	Energy
Zika virus SPH2015	957 -- 990	Selenoprotein P mRNA	554 -- 586	-24.72210

Details of Selected Interaction Download Interaction Details

```
                     956                                 991
                     |                                   |
        5'-UAG...CCUCA        GA  G          A           CACAC...AGA-3'
 Target                GGUU CUU  AU AUAUGGGAAA GUUAGGAAGGA
                       ||:| |||  |: |||||||||| ||||||||||
  Query                CCGA GAA  UG UAUACCCUUU CAAUCCUUCCU
        3'-AAU...AAUUA    A   GA                          UUUUC...AUA-5'
                     |                                   |
                     587                                 553
```

Energy	**-24.72210 kcal/mol**	Position - Target RNA	957 -- 990
Hybridization Energy	-37.9	Position - Query RNA	554 -- 586
Unfolding Energy - Target	7.05878	Position Seed - Target RNA	984 -- 990
Unfolding Energy - Query	6.11916	Position Seed - Query RNA	554 -- 560

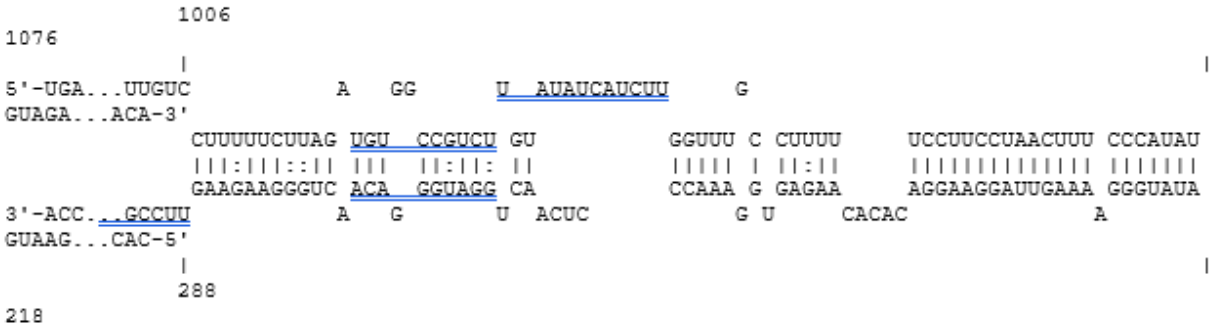
```
Sequence   : CCUCAGGUUCUUGAAUGAUAUGGGAAAAGUUAGGAAGGACACAC&CUUUUUCCUUCCUAACUUUCCCAUAUGUAGAAGAAGCC
Structure  : .....(((((((..((.((((((((((.((((((((((.....&.....)))))))))))))))))))))))..))).))))
StartIndex : 952&549
```

```
IntaRNA default input parameters used for all of the analyses in this
study:

Energy threshold for output                          0
        Number of suboptimal interactions            0
Seed Parameters
        Min. number of basepairs in seed             7
        Max. Number of mismatches in seed            0
        Seed position (query)                        not provided
Folding Parameters
        Temperature for energy computation           37.0
        Folding window size target                   150
        Max. basepair distance (target)              100
```

Figure A2: Raw data output for IntaRNA-predicted RNA-RNA interaction between a 3' region of the ZIKV mRNA (polymerase coding region) and a region of SePP1 mRNA. A more extended version of this interaction, extending into an upstream intron of SePP1, is shown as **Figure 5** in the main article and as raw data below.

Target	Position	Query	Position	Energy
SePP1 unspliced 8001-10000	1007 -- 1075	Zika 9501-10000	219 -- 287	-25.66320



Energy	-25.66320 kcal/mol
Hybridization Energy	-59.6
Unfolding Energy - Target	10.05310
Unfolding Energy - Query	23.92360
Position - Target RNA	1007 -- 1075
Position - Query RNA	219 -- 287
Position Seed - Target RNA	1069 -- 1075
Position Seed - Query RNA	219 - 225

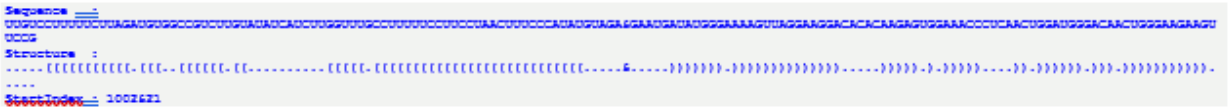


Figure A3: Raw output data for IntaRNA-predicted RNA-RNA interaction between a 3' region of the Zika virus mRNA (polymerase coding region) and a region of SePP1 pre-mRNA, spanning an intron/exon boundary in SePP1. For this analysis, a 500 base fragment of Zika virus (9,501-10,000) was scanned vs. a 2,000 base fragment (8,001-10,000) of the SePP1 genomic sequence. Add 9500 to the numbering shown below for the position in the Zika gene file GenBank KU321639.1, and add 8000 to the numbering shown below for the position in the SePP1 gene file GenBank DQ022288.1. The 2000 base SePP1 Sequence includes the entire 4th exon of only 117 bases (where there is an antisense match between Zika and the SePP1 mRNA coding region), extended to include as much of the upstream and downstream introns as can be included in the 2000 base size limit for IntaRNA input. The same results are shown as Figure 3A in the main article, with the position numbering edited to correspond to the respective GenBank sequences.

Target	Position	Query	Position	Energy
TXNRD1 55001-57000	1097 -- 1132	Zika 3' non-coding sfRNA	36 -- 67	-24.75450

Details of Selected Interaction Download Interaction Details

```
                    1096                                1133
                     |                                   |
          5'-GUU...GGACU   A   UUU               U        AGAAG...CUC-3'
  Target               CUG UAC   GUUUUCCCCAAGCUGU GGCUGAUUA
                       ||| :||   |:|||||||||||||| ||||||:||
   Query               GAC GUG   CGAAAGGGGUUCGACA CCGACUGAU
          3'-AGA...UGUCC       U                           CGUCC...UCG-5'
                     |                                   |
                     68                                  35
```

Energy	**-24.75450 kcal/mol**	Position - Target RNA	1097 -- 1132
Hybridization Energy	-48.0	Position - Query RNA	36 -- 67
Unfolding Energy - Target	7.64143	Position Seed - Target RNA	1126 -- 1132
Unfolding Energy - Query	15.60400	Position Seed - Query RNA	36 -- 42

```
Sequence   : GGACUCUGAUACUUUGUUUUCCCCAAGCUGUUGGCUGAUUAAGAAG&CCUGCUAGUCAGCCACAGCUUGGGGAAAGCUGUGCAGCCUGU
Structure  : .....(((.(((...(((((((((((((((((.(((((((((.....&.....)))))))))))))))))))))))).))))))).....
StartIndex : 1092&31
```

Job resubmission

Use the following button if you want to resubmit the job with altered input or parameterization: RESTART

When using IntaRNA please cite :

- Patrick R. Wright, Jens Georg, Martin Mann, Dragos A. Sorescu, Andreas S. Richter, Steffen Lott, Robert Kleinkauf, Wolfgang R. Hess, and Rolf Backofen
  CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains
  Nucleic Acids Research, 2014, 42 (W1), W119-W123.
- Anke Busch, Andreas S. Richter, and Rolf Backofen
  IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions
  Bioinformatics, 24 no. 24 pp. 2849-56, 2008

Results are computed with IntaRNA version 1.2.5 (wrapper 1.0.7.1) linking Vienna RNA package 1.8.4

Figure A4: Raw data output for IntaRNA-predicted RNA-RNA interaction between a 3' region of the ZIKV non-coding "subgenomic flavivirus RNA" (sfRNA) and an intronic region of the human thioredoxin reductase 1 (TXNRD1) pre-mRNA. The Zika 300 base 3' UTR region was scanned vs. fragment 55001-57000 of the TXNRD1 genomic sequence using the IntaRNA program with default parameters. The same results are shown as Figure 5 in the main article, with the position numbering edited to correspond to the respective GenBank sequences

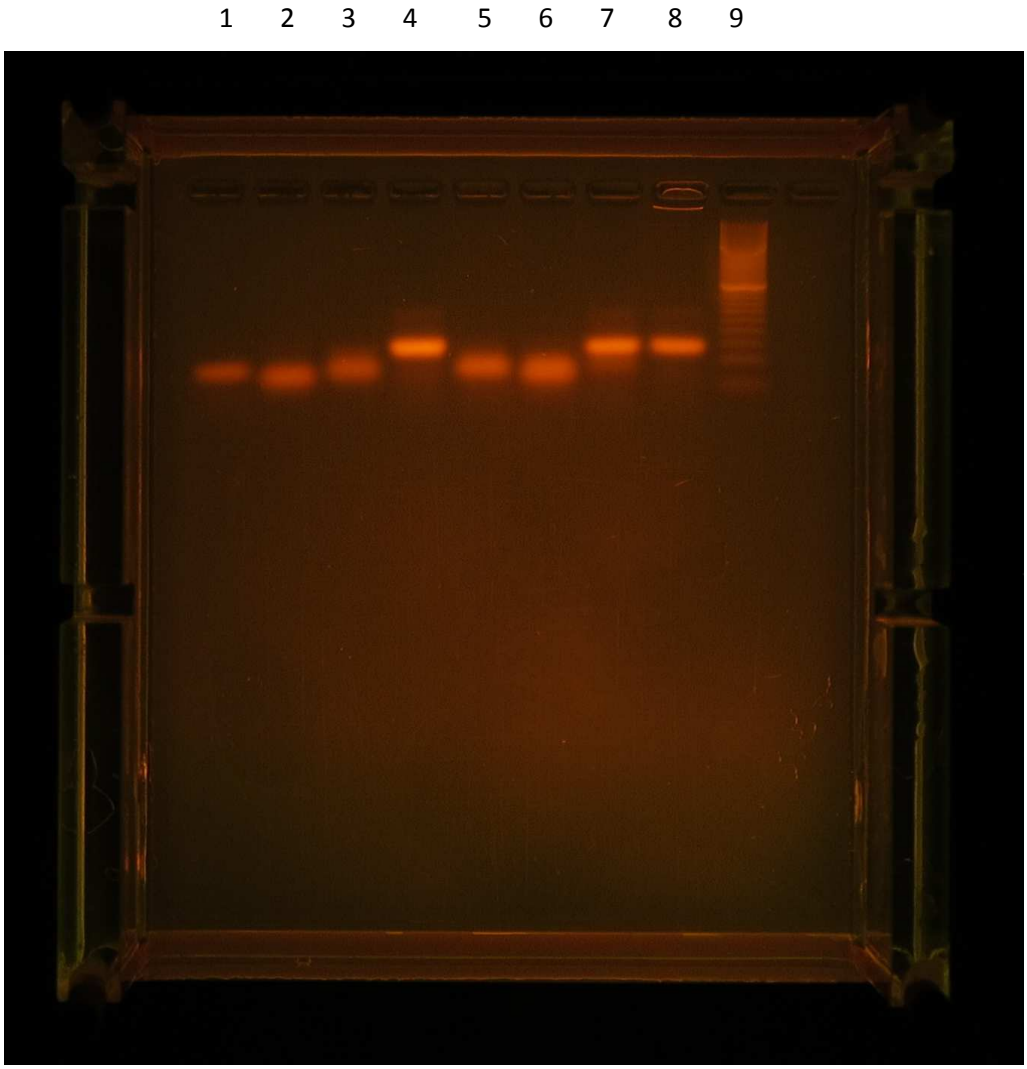


Figure A5: Original uncropped gel photo corresponding to Fig. 6 in the main article, showing target-specific *in vitro* DNA hybridization of the predicted ZIKV-SePP1 antisense pairing of Figure 5, by gel shift assay using DNA oligonucleotides. Lanes 1-6 are as described in the legend to Figure 1B. The additional lanes not included in Figure 1B are as follows, Lane 7: same as lane 4 but with a 2:1 ratio of viral to SePP1 DNA, giving two bands, dsDNA (higher bright band) and unpaired Zika (fainter lower band.). Lane 8: same as lane 4 but in the presence of sheared human cellular DNA. Lane 9: size marker for dsDNA, the relevant markers (from the lowest up) being 20bp, 40bp and 60 bp. The Zika:SePP1 hybrid dsDNA runs above the 40bp and just below the 60bp standard.

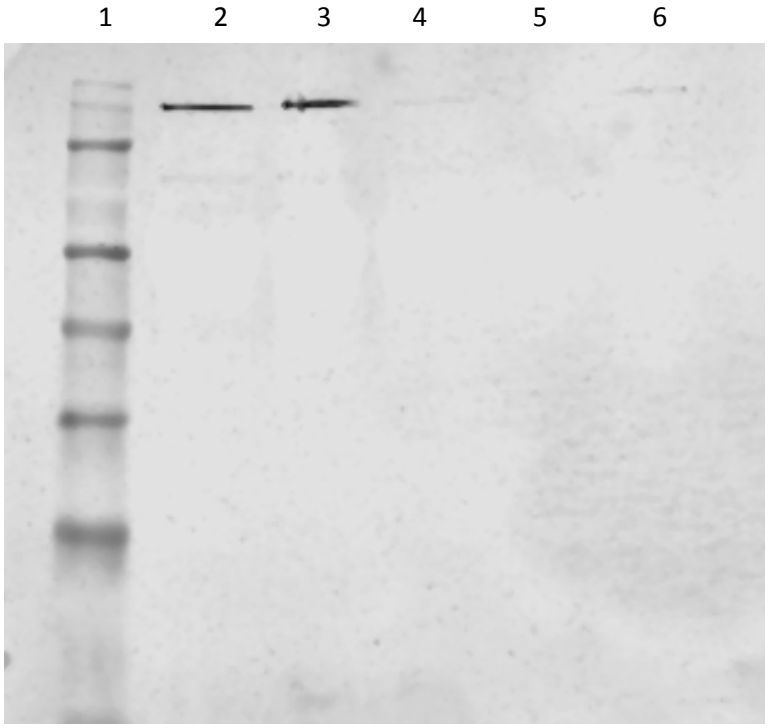


Figure A6: Original uncropped Western blot of SePP1 protein expression, visible at the 60 kDa mass level, under varied conditions corresponding to Figure 7 in the main article. Lanes 2, 3, 4 and 6 correspond to the labeled lanes shown and described in the legend to Figure 4A. Lane 5 below was a skipped lane while loading the gel due to a noticed imperfection in the well, and was cropped out of Figure 4A. Lane 6 corresponds to the HEK+ZIKV+Se sample. All lanes and protein samples were loaded and prepared as described in the Materials and Methods section in the main article as well as the legend to Figure 7. This is one of 3 similar Western blots that were analyzed by densitometry using ImageJ (Table A1).

Table A1: Data analysis of ImageJ densitometry results for SePP1 expression. Top: Densitometry values from Western Blot analysis for the protein expression of SePP1 from each experimental condition. Data gathered from triplicate blots all containing the same amount of protein sample from each experimental condition. P-values generated by comparing SePP1 expression in experimental conditions (HEK+Se, HEK+ZIKV and HEK+ZIKV+Se) to the HEK only condition in a two-sampled, 2-tailed t-Test assuming unequal variance. SEM represents standard error of the mean. Percent change from HEK for each experimental condition was calculated using the average densitometry values for each condition. * indicates a p-value ≤0.05.

Experimental Condition	**HEK**	**HEK+Se**	**HEK+ZIKV**	**HEK+ZIKV+Se**
Blot 1	42,200	63,800	0	0
Blot 2	41,600	123,000	321	394
Blot 3	46,800	53,600	484	854
Average value	43,500	80,100	268	416
SEM	±1,640	±21,600	±155	±240
P-value		0.23	0.0015*	0.0015*
Percent change from HEK		+84.1	-99.4	-99.1

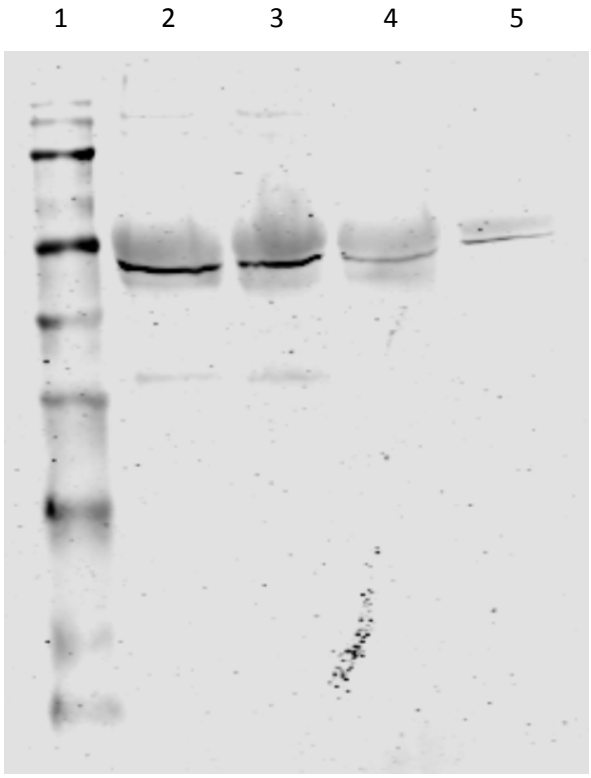


Figure A7: Original uncropped Western blots of TXNRD1 protein expression under varied conditions corresponding to Figure 8 in the main article. Lanes 2-5 below correspond to the labeled lanes shown in Fig. 8A. All lanes and protein samples were loaded and prepared as described in the Materials and Methods section in the main article as well as the legend to Figure 8A. This is one of 3 similar Western blots that were analyzed by densitometry using ImageJ (Table A2).

Table A2: Data analysis of ImageJ densitometry results for TXNRD1 expression. Top: Densitometry values from Western Blot analysis for the protein expression of TXNRD1 from each experimental condition. Data gathered from triplicate blots all containing the same amount of protein sample from each experimental condition. P-values generated by comparing TXNRD1 expression in experimental conditions (HEK+Se, HEK+ZIKV and HEK+ZIKV+Se) to the HEK only condition in a two-sampled, 2-tailed t-Test assuming unequal variance. SEM represents standard error of the mean. Percent change from HEK for each experimental condition was calculated using the average densitometry values for each condition. * indicates a p-value ≤0.05.

Experimental condition	**HEK**	**HEK+Se**	**HEK+ZIKV**	**HEK+ZIKV+Se**
Blot 1	55,800	40,600	7,290	8,350
Blot 2	52,800	23,000	3,220	5,230
Blot 3	30,900	21,200	2,940	4,510
Average	46,500	28,300	4,480	6,030
SEM	±7,850	±6,190	±1,410	±1,180
P-value		0.14	0.034*	0.036*
Percent change from HEK		-39.1	-90.4	-87.0

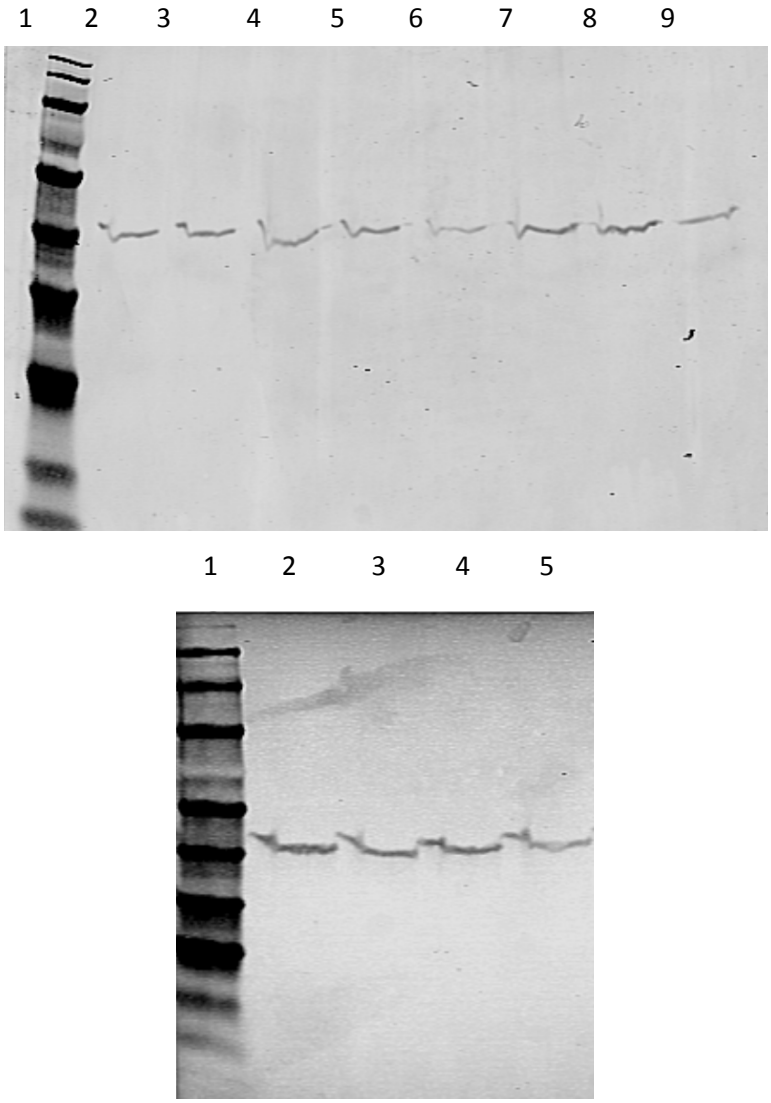


Figure A8: Uncropped Western blots of Beta-actin protein expression under varied conditions in triplicate. Both top panel and bottom panel blots are identical containing equal amounts of protein, ran under the same conditions, and containing the same amount of molecular weight marker standard control. All samples and blots were run identically as described in the Materials and Methods section of the main article. Lanes are assigned for the top panel as follows: Ln1: molecular weight marker, Ln2: HEK, Ln3: HEK+Se, Ln4: HEK+ZIKV, Ln5: HEK+ZIKV+Se, Ln6: HEK, Ln7: HEK+Se, Ln8: HEK+ZIKV, Ln9: HEK+ZIKV+Se. Lanes are assigned for bottom panel as follows: Ln1: molecular weight marker, Ln2: HEK, Ln3: HEK+Se, Ln4: HEK+ZIKV, Ln5: HEK+ZIKV+Se. Bars corresponding to the molecular weight of Beta-actin at the 40 kdal line are observed for each sample consistently.

Table A3: Data analysis of ImageJ densitometry results for Beta-actin expression. Top: Densitometry values of Beta-actin expression across all experimental conditions. Data gathered from triplicate blots all containing the same amount of protein sample from each experimental condition. P-values generated by comparing Beta-actin expression in experimental conditions (HEK+Se, HEK+ZIKV and HEK+ZIKV+Se) to the HEK only condition using a two sampled, 2-tail t-Test assuming unequal variences. Bottom: Densitometry values from western blot analysis for the total protein expression of Beta-actin from each western blot. Data gathered from triplicate blots all containing the same amount of protein sample from each experimental condition. Western blot analysis of Beta-actin was performed as a western blot control to ensure the western blot protocol was effective and experimental results concerning SePP1 and TXNRD1 protein expression were a result of ZIKV infection and not a variable concerning the western blot techniques utilized in these experiments. No meaningful changes in Beta-actin expression occurred in any experimental condition. SEM represents standard error of the mean.

Experimental condition	**HEK**	**HEK+Se**	**HEK+ZIKV**	**HEK+ZIKV+Se**
Blot 1	758	804	1,040	1,310
Blot 2	1,050	759	1,580	1,560
Blot 3	1,500	991	2,060	716
Average value	1,100	851	1,560	1,120
SEM	±216	±71.1	±295	±251
P-value		0.38	0.28	0.79

Blot number	**Total average Beta-actin expression for blot**	**SEM**
Blot 1	978	±127
Blot 2	1,240	±201
Blot 3	1,320	±296

Table A4: RT-qPCR data analysis in quadruplicate for each experimental condition for SePP1, TXNRD1, and ZIKV RNA levels. Lower Ct values indicate more starting material in the original RNA sample. All p-values were calculated using a two sampled, 2-tailed t-Test assuming equal variance. P-values in the 8th column were generated by comparing the quadruplicate Ct values for each experimental condition against the quadruplicate Ct values for the HEK only condition. P-values in the 9th column generated by comparing all Ct values for the HEK+Se experimental condition against the all Ct values for the HEK+ZIKV+Se condition. P-values in 10th column generated by comparing all Ct values for the HEK+ZIKV experimental condition against the all Ct values for the HEK+ZIKV+Se experimental condition. All Ct values are ROX normalized.

Experiment condition	Run 1	Run 2	Run 3	Run 4	Ct average	Standard deviation
SePP1 RNA levels						
HEK	22.16	19.22	24.28	16.65	20.58	±3.3
HEK+ZIKV	17.74	18.52	17.37	18.33	17.99	±0.53
HEK+Se	21.02	20.23	19.06	19.29	19.90	±0.90
HEK+ZIKV+Se	17.91	19.63	19.66	19.34	19.14	±0.83
TXNRD1 RNA levels						
HEK	17.24	19.44	22.57	19.96	19.80	±2.2
HEK+ZIKV	14.61	20.39	23.19	23.06	20.31	±4.0
HEK+Se	18.4	17.14	20.35	21.02	19.23	±1.8
HEK+ZIKV+Se	14.72	17.62	19.40	15.48	16.81	±2.1
ZIKV RNA levels						
HEK	29.15	29.03	27.09	27.22	28.12	±1.1
HEK+ZIKV	26.18	25.43	25.57	25.17	25.59	±2.9
HEK+Se	28.17	30.42	28.50	23.69	27.70	±0.43
HEK+ZIKV+Se	26.03	25.78	29.98	25.54	26.83	±2.1

APPENDIX B:

SUPPLEMENTAL FIGURES AND TABLES FOR CHAPTER III

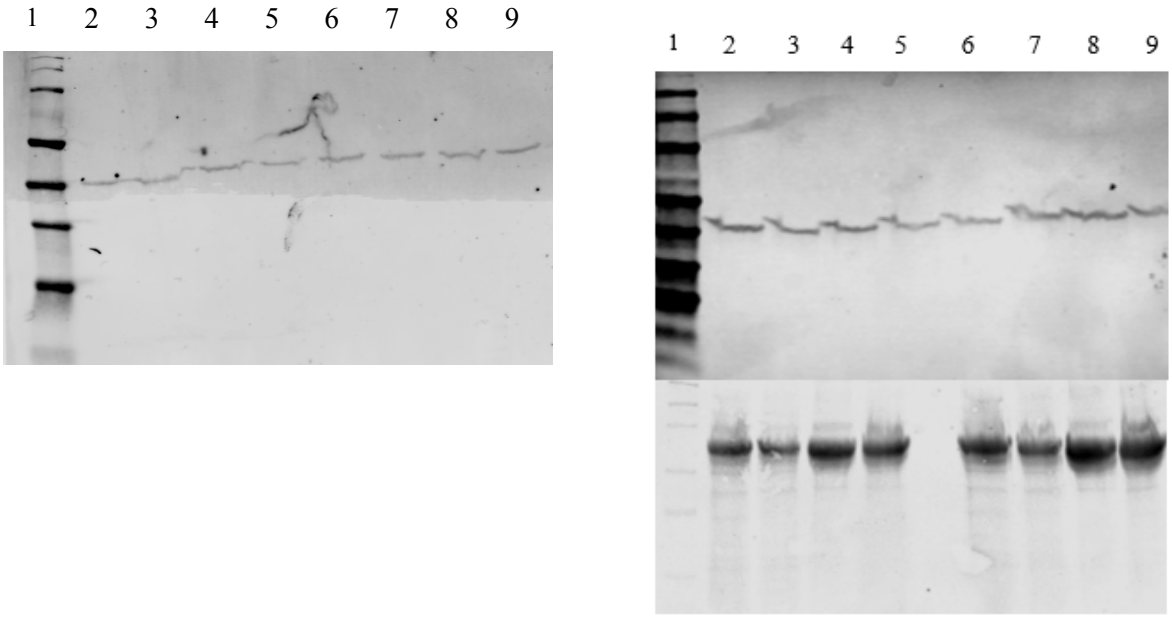


Figure B1: Beta-actin expression across all experimental conditions in triplicate. Beta-actin western blot probe was performed as a control to eliminate any variable in the execution of the experimental western blot probe for nef-GFP expression. All gels were prepared in the same way. Equal amounts of protein were loaded into each lane with 3 uL 10X loading dye and brought up to 30 uL with 1x PBS. Ln1: 2 uL molecular weight marker, ln2: HEK lysate, ln3: HEK lysate+Se, ln4: pnefStop, ln5: pnefStop+Se, ln6: pnefATI1, ln7: pnefATI1+Se, ln8:PAC, ln9: PAC+Se. Gels were imaged with an Odyssey CLx (LI-COR) gel imager excited at 700 nm with emission measured at 720 nm with 84uM concentration on the highest quality setting. Top panel has a foreign object on the image which was an embedded fiber that merged to the nitrocellulose membrane during transfer but has no effect on the outcome of the western blots presented here. In bottom, a lane was skipped between lanes 5 and 6 due to an inconsistency in the well from the package. All beta-Actin expression comes to the 45 kdal line which is what is expected and no changes in expression of Beta-actin are observed. No meaningful change in Beta-actin expression was observed across the experimental conditions.

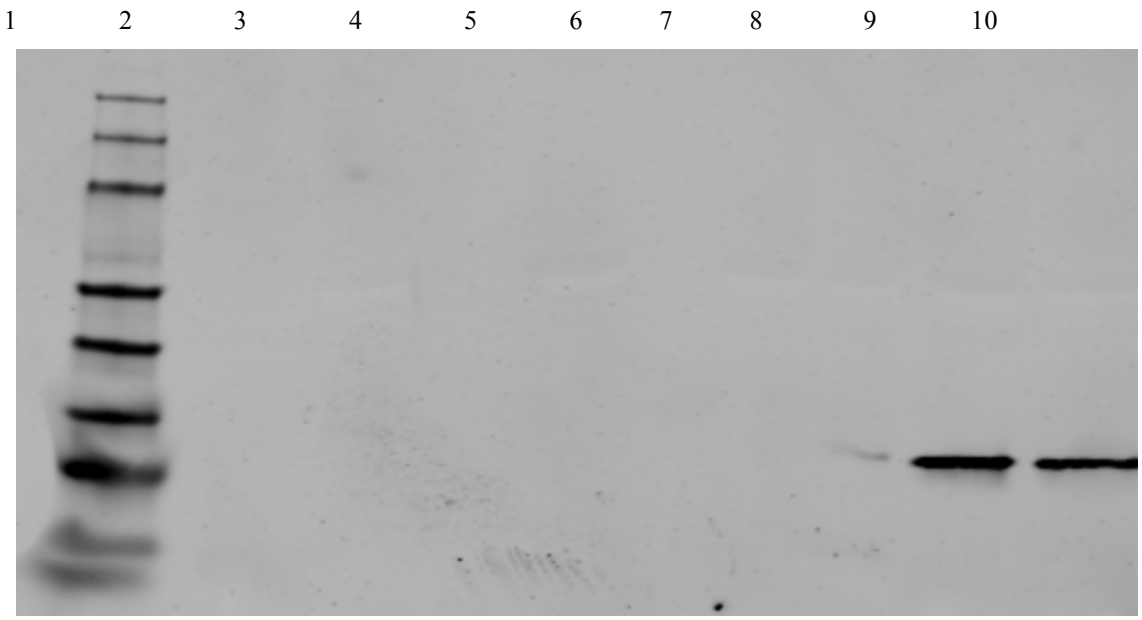


Figure B2: GFP expression across all experimental conditions. All gels were loaded identically and are as follows: Equal amounts of protein were loaded into each lane with 3 uL 10X loading dye and brought up to 30 uL with 1x PBS. Ln1: 2 uL molecular weight marker, ln2: HEK lysate, ln3: HEK lysate+Se, ln4: pnefStop, ln5: pnefStop+Se, ln6: pnefATI1, ln7: pnefATI1+Se, ln8:PAC, ln9: PAC+Se. In Panels A and C, a lane was skipped between the molecular weight ladder (ln1) and the HEK lysate (ln22) due to pipetting error. In Panel C, ln7: pnefATI1 and ln6: pnefATI1+Se. Gels were imaged with an Odyssey CLx (LI-COR) gel imager excited at 700 nm with emission measured at 720 nm with 84uM concentration on the highest quality setting. All GFP expression comes to the 27 kdal line which is what is expected for GFP alone but the fused nef-GFP protein product was expected to have more of a gel upshift than what was observed for all the western blot probes.

Table B1: Image analysis values from western blot triplicates for the protein expression of GFP from each experimental condition. Data gathered from triplicate blots all containing the same amount of protein sample from each experimental condition. All samples for each experimental condition were generated from the same cell condition. Gels were scanned with an Odyssey CLx (LI-COR) gel imager excited at 700 nm with emission measured at 720 nm with 84uM concentration on the highest quality setting. Quantitative analysis of GFP expression was also performed using the Odyssey CLx (LI-COR) gel imager analysis software. P-values generated by comparing GFP expression in experimental conditions without selenium supplementation to experimental conditions with selenium supplementation. All p-values were generated by in a two-sampled, 2-tailed t-Test assuming unequal variance. * indicates a p-value ≤0.05.

Experimental Condition	Blot 1	Blot 2	Blot 3	Average	Standard Deviation	P-value
HEK	0	0	0	0	0	
HEK+Se	0	0	0	0	0	
pnefStop	0	0	0	0	0	
pnefStop+Se	0	0	0	0	0	
pnefATI1	0	0	0	0	0	
pnefATI1+Se	973	898	1,880	1,250	±547	0.012*
PAC	90,600	64,600	93,100	82,700	±15,800	
PAC+Se	58,300	62,300	58,100	59,500	±2,370	

APPENDIX C:

SUPPLEMENTAL FIGURES AND TABLES FOR CHAPTER IV

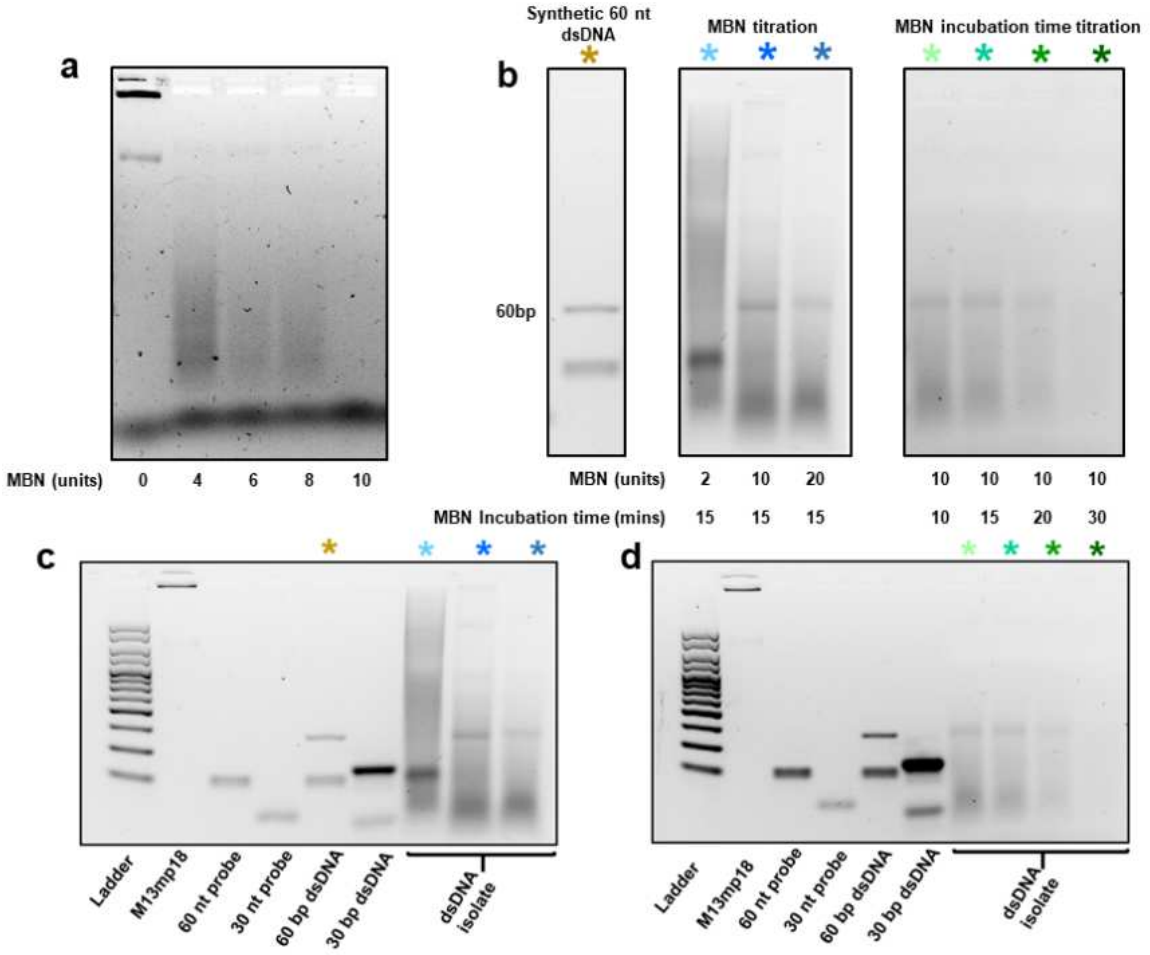


Figure C1: Gel optimization of amount and incubation time for Mung Bean Nuclease (MBN) digestion of single-strand M13mp18 DNA. (a) Digestion of 1.25 µg M13mp18 in the presence of different amounts of MBN as indicated. 10 units of MBN is required to fully digest the material. (b) Isolation of 60 bp dsDNA product from M13mp18 with MBN. Both the amount of MBN and incubation time were varied. M13mp18 annealed with 60 nt probe requires 10 units of MBN and 15 mins of incubation at 30 o C to digest single strand regions. Full gels are shown in (c) and (d). Asterisk colors indicate lanes shown in (b).

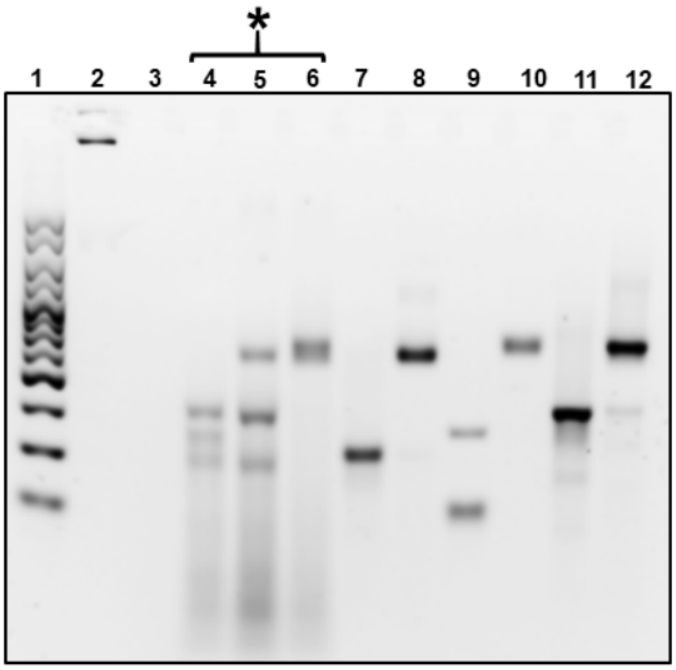


Figure C2: Full gel showing isolation of multiple distinct products from M13mp18. Lane 1: ladder; lane 2: single-strand M13mp18; lane 3: M13mp18 fully digested by MBN; lane 4: isolated biotinylated products from M13mp18 using three DNA probes (50, 60, and 75 nt); lane 5: electromobility shift assay in the presence of MS when only the 60 nt probe is biotinylated; lane 6: electromobility shift assay in the presence of MS when all probes are biotinylated. Lanes 7, 9, and 11 show synthetic 50 nt dsDNA, 60 nt dsDNA, and 75 nt dsDNA, respectively. Their electromobility shift assays are shown in lanes 8, 10, and 12, respectively. Lanes marked with an asterisk are shown in **Figure 22a** of the main text.

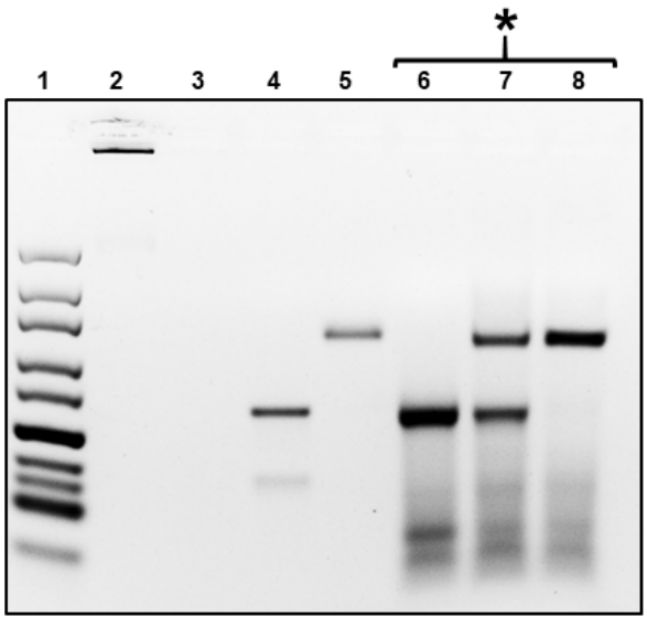


Figure C3: Full gel showing isolation of two products of the same length from M13mp18. Lane 1: ladder; lane 2: single-strand M13mp18; lane 3: M13mp18 digested by MBN; lane 4: synthetic monobiotinylated 60 bp duplex construct; lane 5: synthetic monobiotinylated 60 bp duplex construct bound to MS; lane 6: two independent biotinylated 60 bp isolates; lane 7: electromobility shift assay in the presence of MS when only one 60 nt probe is biotinylated; lane 8: electromobility shift assay in the presence of MS when both 60 nt probes are biotinylated. The lanes marked with an asterisk are shown in **Figure 20b**, inset in the main text.

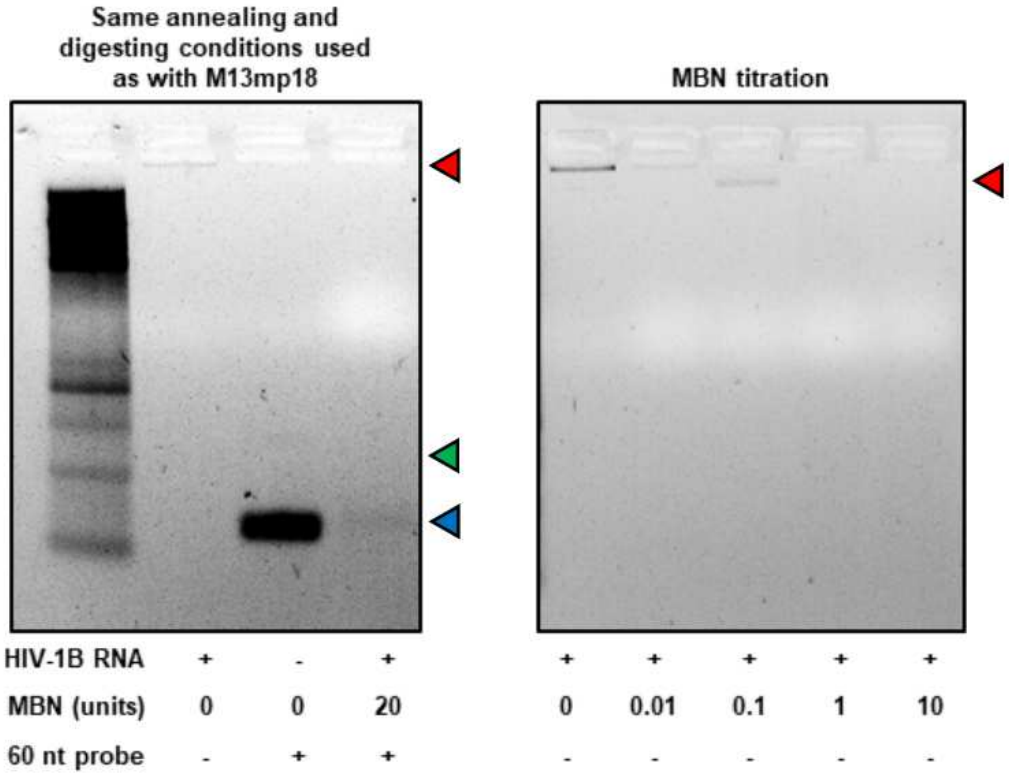


Figure C4: Gel analysis of target isolation from full-length HIV-1B RNA. Left: isolation carried out under identical conditions used to isolate 60 bp target from M13mp18. Lane contents indicated at bottom (lane one shows a ladder). The product (anticipated position marked with green arrow) was digested along with unannealed probe (blue arrow) and single strand HIV-1B (red arrow). Right: MBN titration (amounts indicated) carried out against 10 µg HIV-1B RNA. Undigested RNA (red arrow) was observed below 1 U.

Table C1: List of probes. In each sequence, T represents biotinylated thymine.

Probes	Complimentary (M13mp18/HIV-1B)	Sequences (5' ⟶ 3')
50 nt	M13mp18	AAGGTAAAGTAATTCTGTCCAGACGACGACAATAAACAACATGTTCAGCT
60 nt -1	M13mp18	CGAACTAACGGAACAACATTATTACAGGTAGAAAGATTCATCAGTTGAGATTTAGGAATA
60 nt - 2	M13mp18	ACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGC
75 nt	M13mp18	TGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTGCA TGCCTGCAGGTC
60 nt	HIV-1B	ACAGTCTACTTGTCCATGCATGGCTTCTCCTTTTAGCTGACATTTATCACAGCTGGCTAC

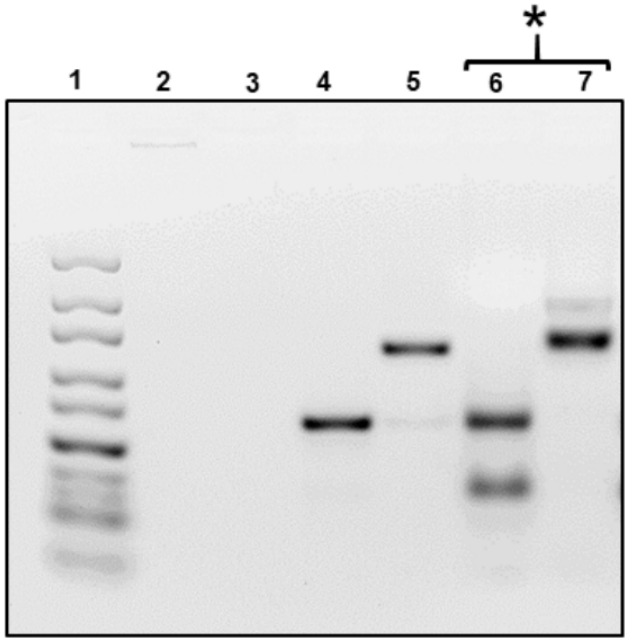


Figure C5: Gel analysis of sequence isolation from full-length HIV-1B RNA. Lane 1: ladder; lane 2: HIV-1B alone; lane 3: HIV-1B RNA after MBN digestion; lane 4: synthetic monobiotinylated 60 bp DNA construct, Lane 5: synthetic monobiotinylated 60 bp DNA construct bound to MS; lane 6: biotinylated 60 bp RNA/DNA heteroduplex isolated from full length HIV-1B RNA; and lane 7: biotinylated 60 bp RNA/DNA heteroduplex isolated from full length HIV-1B RNA bound to MS. The lanes marked with an asterisk are shown in **Figure 21a** in the main text.

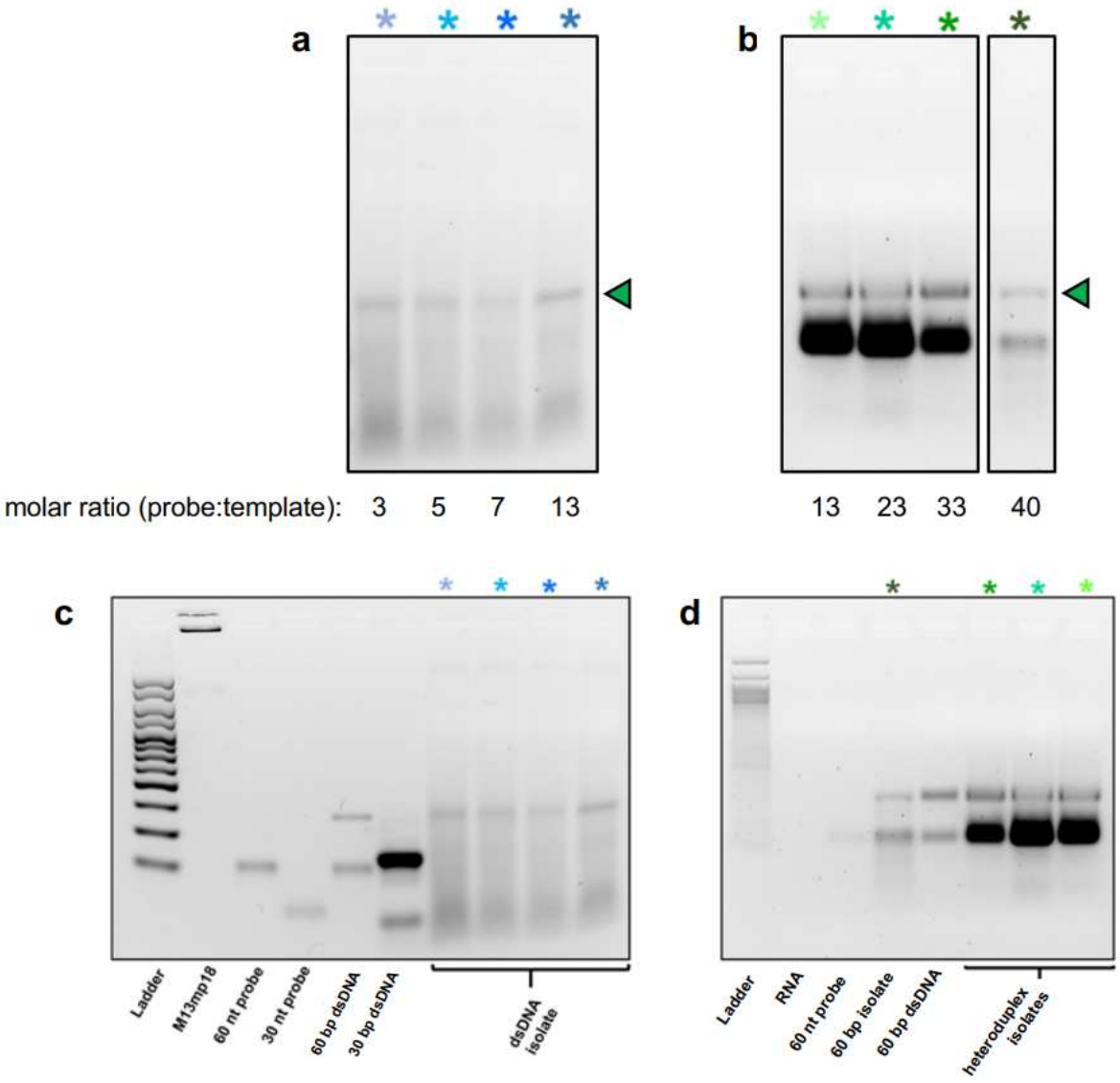


Figure C6: Gel optimization of probe required for efficient annealing to both M13mp18 DNA. (a) and HIV-1B RNA (b). Gels are cropped for easy comparison. Molar ratios indicated at bottom. Increasing amount of 60 nt probe yields higher relative intensity of the 60 bp dsDNA or RNA/DNA heteroduplex product (green arrows). The original gels are shown in (c) and (d) and the colors of asterisks indicate lanes shown in (a) and (b).